Ph.D.Dissertation

Design of High-Speed CMOS Interface Circuits for Optical Communications

광통신을 위한 고속 CMOS 인터페이스 회로설계

by

Gyu-Seob Jeong

August, 2017

School of Electrical Engineering and Computer Science
College of Engineering
Seoul National University

Design of High-Speed CMOS Interface Circuits for Optical Communications

지도 교수 정 덕 균

이 논문을 공학박사 학위논문으로 제출함

2017 년 8 월

서울대학교 대학원

전기·컴퓨터공학부

정 규 섭

정규섭의 박사 학위논문을 인준함

2017 년 8 월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Design of High-Speed CMOS Interface Circuits for Optical Communications

by

Gyu-Seob Jeong

A Dissertation Submitted to the Department of
Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy
at
SEOUL NATIONAL UNIVERSITY

August, 2017

Committee in Charge:

Professor Hyuk-Jae Lee, Chairman

Professor Deog-Kyoon Jeong, Vice-Chairman

Professor Suhwan Kim

Professor Yongsam Moon

Professor Jung-Hoon Chun

Abstract

The bandwidth requirement of wireline communications has increased exponentially because of the ever-increasing demand for data centers and high-performance computing systems. However, it becomes difficult to satisfy the requirement with legacy electrical links which suffer from frequency-dependent losses due to skin effect, dielectric loss, channel reflections, and crosstalk, resulting in a severe bandwidth limitation. In order to overcome this challenge, it is necessary to introduce optical communication technology, which has been mainly used for long-reach communications, such as long-haul networks and metropolitan area networks, to the medium- and short-reach communication systems. However, there still remain important issues to be resolved to facilitate the adoption of the optical technologies. The most critical challenges are the energy efficiency and the cost competitiveness as compared to the legacy copper-based electrical communications. One possible solution is silicon photonics that has long been investigated by a number of research groups. Despite inherent incompatibility of silicon with the photonic world, silicon photonics is promising and is the only solution that can leverage the mature CMOS technologies.

In this thesis, we summarize the current status of silicon photonics and provide the prospect of the optical interconnection. We also present key circuit techniques essential to the implementation of high-speed and low-power optical receivers. And then, we propose optical receiver architectures satisfying the aforementioned requirements with novel circuit techniques.

Keywords : CMOS technologies, optical communication, optical receiver, silicon photonics

Student Number : 2012-20856

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Today, people's everyday lives are connected online, meaning that anyone can access or share data worldwide at any time and at any place. For example, most of the people today are familiar with accessing streaming media from YouTube, using social network services, and cloud services for personal purposes. The concept of internet of things (IoT) is not new anymore, as the IoT devices have already become a part of people's lives. Two important factors contributing to this revolution are the developments of high-performance computing systems and data communication systems which are closely interdependent. Especially, the requirement of high-performance data communication systems is being emphasized than ever, because of ever-increasing demand for data throughput in every computing system. Even a

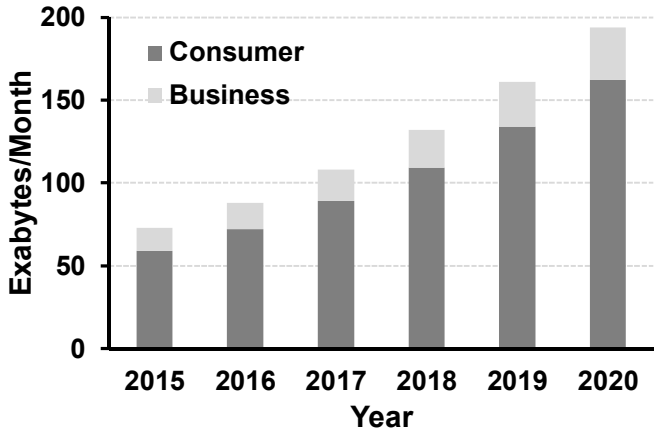
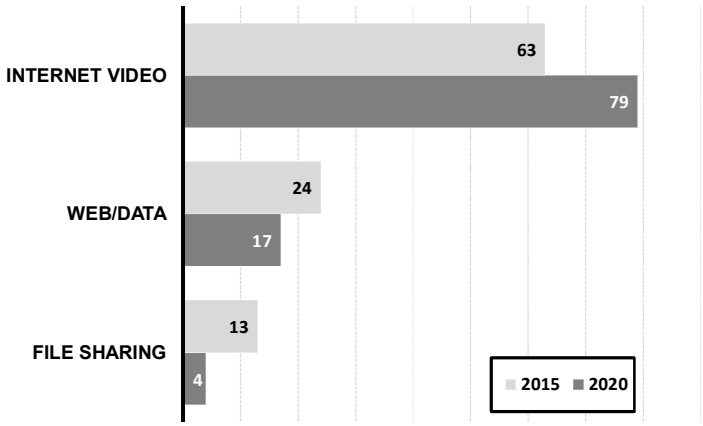


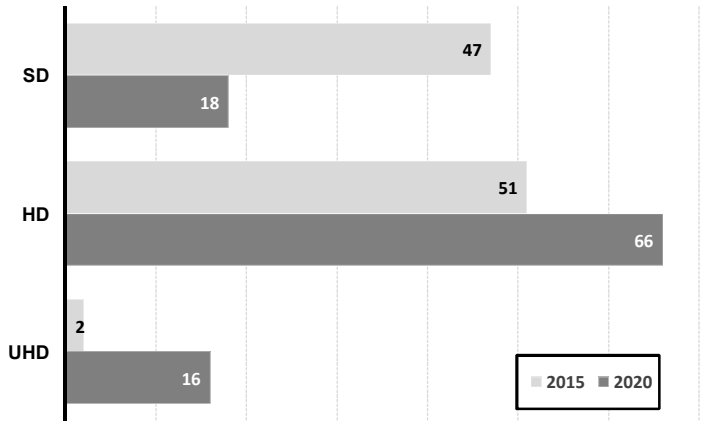
Fig. 1.1 Forecast of total global IP traffic.

microprocessor and memory constituting a tiny computing system is now required to handle data rates of several hundreds of Gb/s, and the data rates are increasing at a relentless rate. A similar phenomenon is observed in longer-distance applications such as inter-server and long-haul communications. The forecast of data throughput demands is provided by Cisco, one of the leading companies specialized in networking equipment. Figure 1.1 forecasts the annual global IP traffics by 2020 with a compound annual growth of 20 %, predicting that the IP traffic would become 200 Exabytes/month in 2020 [1]. It is worth investigating the main factor that results in these explosive demands of total data throughput.

Figure 1.2 illustrates the breakdown of the total global IP traffic by application. Note that, in Figure 1.2(a), the only increase is observed in internet video applications. This is reasonably explained in Figure 1.2(b) by further subdividing video applications into SD, HD, and UHD. From these investigations, it is evident that high-



(a)



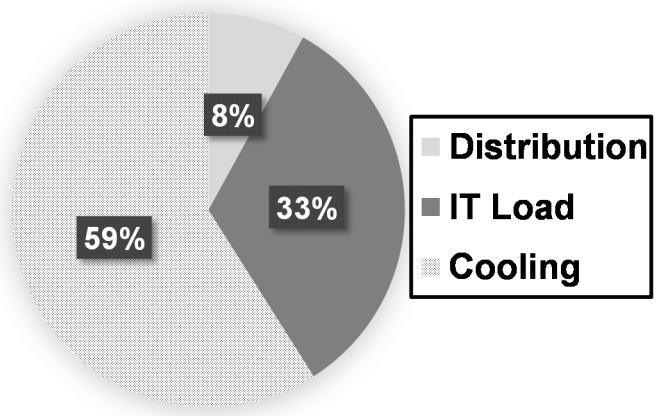
(b)

Fig. 1.2 Forecast of global IP traffic (a) by application and (b) by video content.

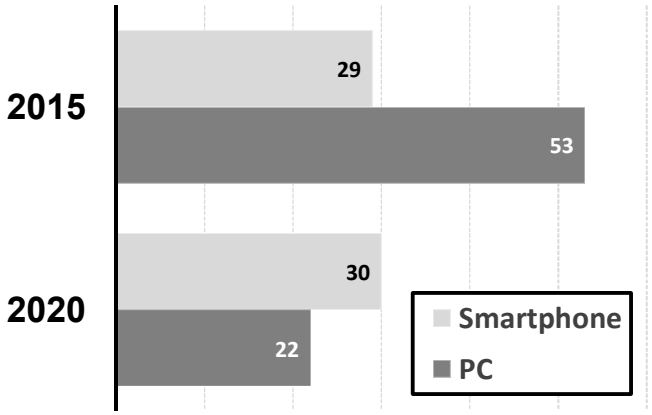
quality video contents would be dominant in the future IP traffic because of increasing demands for 3D media contents and virtual reality. It can also be predicted that the capability of handling high data bandwidth will be required for not only data

centers but also end-user devices.

Another important requirement is the reduction of power consumption. Figure 1.3(a) shows the power breakdown of typical data centers, indicating that the current communication systems are consuming much power in handling extremely high data



(a)



(b)

Fig. 1.3 (a) Power breakdown of data centers and (b) IP traffic contribution.

rates so that a significant part of resources is being wasted for cooling systems. In the case of end-user devices, the proportion of mobile devices will exceed that of PCs as shown in Figure 1.3(b), meaning the importance of power management will be undoubtedly more pronounced in the future. In summary, it is required to build a reliable communication system satisfying the aforementioned requirements, necessitating innovations in the existing communication systems.

The primary obstacle that hinders high-speed operation is the electrical limitation of the copper-based interconnection which is predominantly used in most applications except for long-reach interconnections. The frequency response of a copper channel severely degrades due to conductor loss and dielectric loss at higher frequencies. Channel reflection and crosstalk are also detrimental to high-speed operation, limiting the maximum operating frequency to only several GHz over a few tens of meters. Even if various circuit techniques have successfully overcome this physical limit, copper-based interconnection faces the challenge of bandwidth limit as the data rate exceeds several tens of Gb/s. On the other hand, optical interconnection can provide a much higher bandwidth and is free from reflection and crosstalk. In practice, optical fibers have successfully replaced the legacy copper cables over long distances and the extension of their usage into shorter distances appears promising. Furthermore, with advanced silicon photonic technologies, it is anticipated that optics can be used even for micro-scale interconnections. Although several problems remain to be solved, many research groups are working toward the same goal, i.e., the realization of high-performance computing and communication systems on a single chip. Silicon photonics with mature complementary metal-oxide-semiconductor (CMOS) IC technologies will provide solutions to fulfill the afore-

mentioned industrial demands.

1.2 Thesis Organization

In this thesis, we summarize the current status of silicon photonics and provide the prospect of the optical interconnection. We also present key circuit techniques essential to the implementation of high-speed and low-power optical receivers.

This thesis is organized as follows. In Chapter 2, an overview of optical communication is presented, through which we provide a vision and a motivation of our work. In terms of integration, a hybrid integration as well as a monolithic integration is needed for high performance optical communication systems. Thus, the hybrid integration will also be discussed briefly. We also deal with the basis of photodiodes (PD), which will be a prerequisite for understanding the optical receiver. In Chapter 3, before moving on to the main subject of this thesis, we briefly introduce basic circuit topologies and techniques required for designing the optical receiver. In Chapter 4, a low-power 20-Gb/s optical receiver front-end is presented. We focus on circuit techniques employed in this design to achieve low-power and high-speed operations. In chapter 5, a 10-40 Gb/s optical receiver front-end with a bandwidth- and power-scaling function is presented. We focus primarily on the scalability that is essential to energy-efficient link systems. Chapter 6 summarizes the proposed works and concludes this thesis.

Chapter 2

Background of Optical Communication

2.1 Overview of Optical Link

Links or I/O circuits have experienced remarkable advancements and played a major role in the development of data communication systems. However, the bandwidth limitation of copper-based links was already predicted a long time ago and engineers have attempted to determine alternative ways to handle the ever-increasing bandwidth requirement. The optical link based on silicon photonics continues to be the best solution to successfully replace the conventional electrical link [2]. Accordingly, many research groups at not only the leading companies such as Intel and IBM but also universities and national institutions have focused on devel-

oping silicon photonic devices including both the optical and the electrical parts [3], [4]. The results of massive researches over several decades appear to be successful and some remarkable results are already applicable to the industrial world [5]. Optical links are gradually replacing their electrical counterparts in long-haul and high-speed applications. Moreover, the ultimate goal of silicon photonics is defined and envisioned as a “macro chip” [6] or an “on-chip server” [7], indicating that the optical links may hopefully substitute most of the electrical links even on an inter-/intra-chip scale. Nevertheless, it appears that the industry is hesitating to adopt silicon photonics as the primary interconnection method for some practical reasons.

In the case of very short-reach applications ranging from several centimeters to meters, the electrical links are still dominant despite the bandwidth limit of copper-based interconnection. The proliferation of copper-based links in this area can be mainly attributed to the advancements of CMOS technologies and circuit techniques. By utilizing the advanced CMOS technologies and the equalization techniques, operations at higher than 25 Gb/s even in severe loss conditions of more than 40 dB were achieved [8]–[10]. Recently, circuit designers have attempted to overcome the limitations of copper by introducing a pulse-amplitude-modulation (PAM) signaling that can enhance the effective data rate for the same loss condition as the conventional binary signaling [11]–[14]. Figure 2.1 summarizes the binary and the PAM-4 transceivers which recorded the fastest operating speed each year over the past 10 years. Notably, the electrical links face the challenge of bandwidth limitation and the maximum available speed would saturate in the immediate future. Therefore, designers are focusing on developing PAM-4 transceivers in the recent years. However, despite the innovations in the electrical links, one can expect that the capacity

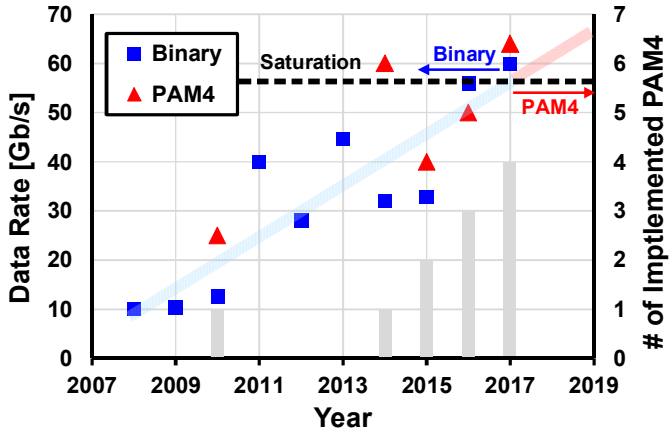


Fig. 2.1 Trends of copper-based electrical links from recent 10-year ISSCC papers.

of the copper-based links would saturate eventually because of both the circuit and the device limitations.

The optical links can be configured with either a multi-mode fiber (MMF) or a single-mode fiber (SMF). An MMF-based link features a distinctly low cost whereas the communication distance is strictly restricted due to modal dispersion. A vertical-cavity surface-emitting laser (VCSEL) is commonly used as a light source for the MMF-based links with a typical wavelength of 850 nm. Owing to the low cost of the MMF-based links, short-reach applications ranging up to several hundreds of meters mostly rely on them. On the other hand, an SMF-based link is free from modal dispersion, thereby providing a much longer communication distance and is only limited by chromatic dispersion and other loss mechanisms such as Rayleigh scattering or electronic absorption. Typical communication standards based on the SMF support a communication distance of up to several tens of kilometers with

wavelengths of 1310 or 1550 nm. Figure 2.2 summarizes communication trends from Ethernet standards: 10 GbE, 40 GbE, and 100 GbE [15], [16]. Among various optical communication applications, the short-reach applications are the most popular and the market size is growing rapidly owing to increasing demand in data centers. Especially, the demand for the active optical cables would increase significantly, as their usage is not restricted to data centers, but can be extended to everyday life. However, since silicon photonics is based on a silicon waveguide, only the telecom wavelengths are usable in silicon photonics, making silicon photonics not directly compatible with the short-reach applications, where extremely low-cost VCSEL-based solutions are preferred. Nevertheless, the prospect of silicon photonics is still promising for the future inter-/intra-chip interconnections and the next-generation data centers covering communication distances of up to 2 km as highlighted in Fig. 2.2 [17], [18].

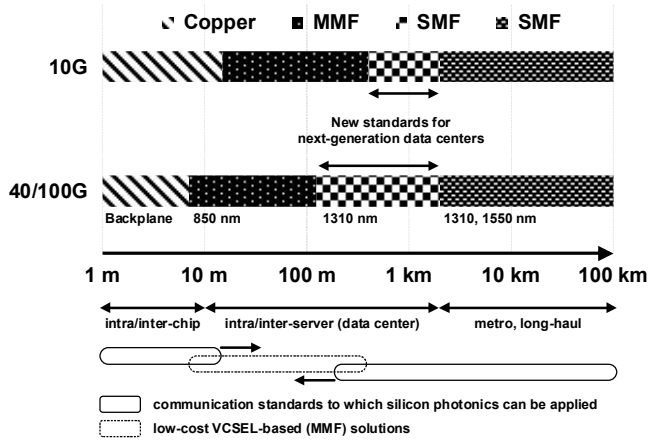


Fig. 2.2 Summary of communication standards, Ethernet.

2.2 Silicon Photonics

In this section, an overview of silicon photonics is presented and silicon photonic technologies are discussed with some examples of silicon photonics transceivers. The current status of silicon photonics and the underlying problems to be solved in the future are addressed.

As explained in [19] and [20], outstanding progresses in silicon photonics were first observed in the early 2000s. Although silicon photonics pursues the realization of on-chip optical interconnections, it may find an immediate application in fiber-optic communications as discussed in the previous section. Therefore, extensive researches on silicon photonics include not only lasers, PDs, modulators, and waveguides but also waveguide-to-fiber couplers. Actually, all the photonic devices necessary to build a complete photonic transceiver were already viable in 2000s except for the laser source. A more recent review work states that the implementation of silicon lasers is still challenging; however, a number of innovations using bonding and epitaxial growth of III-V materials on silicon have been accomplished, so called a hybrid silicon laser [21]. In the case of PDs, silicon by itself cannot be a good absorber for telecommunication wavelengths due to the large bandgap energy. Among the numerous silicon-based PDs, a germanium-introduced PD is the most popular and recent germanium PDs have achieved very high bandwidths of several tens of GHz [22], [23]. Silicon-based PDs will be discussed again in the following section. Silicon modulators are typically based on plasma dispersion effect that the refractive index of the waveguide is changed by the density of free carriers, thereby enabling

an amplitude modulation with Mach-Zehnder interferometer (MZI) structures. Modern silicon MZI modulators are capable of handling the data rates higher than 50 Gb/s [24], [25]. Another type of silicon modulators is a ring resonator which exhibits a small footprint and better energy efficiency than the MZI counterparts [26], [27]. Furthermore, a wavelength-division-multiplexing (WDM) is readily applicable to the ring-resonator-based architectures, which is a desirable feature for future silicon photonic transceivers [28]. Much progress has also been made in silicon waveguides and couplers. Depending on the waveguide type, typical propagation loss can be several dB/cm or less than 1 dB/cm [21]. A coupling technique is also of great importance for successfully interfacing with an off-chip optical fiber. Fortunately, a number of good solutions such as holographic lens have been proposed so far, ideally exhibiting an extremely low-loss condition which is comparable to that of the conventional fiber connector [5].

Despite the brilliant success of the silicon photonic ICs (PICs), compatibility with electronic ICs (EICs) or a monolithic integration of the PICs and the EICs is still challenging. Despite lots effort to realize electronic and photonic ICs (EPICs) on a bulk CMOS platform, the overall performance has yet to be improved. Alternatively, a silicon-on-insulator (SOI) CMOS platform is generally accepted as an ideal platform for the implementation of both silicon waveguides and active transistors.

In a standard 0.13-μm CMOS SOI platform, the first fully integrated optoelectronic transceiver was realized by Luxtera in 2006 [29]. The implemented chip achieves a total throughput of 20 Gb/s with a dual-channel architecture using an SMF for long-reach communication. The overall block diagram of the transceiver is described in Figure 2.3. The TX converts an electrical signal into a data-modulated optical signal and the RX operates in the other way around. The TX equalizer in the front compensates for the frequency loss of the electrical channel prior to the TX input. Following the equalizer, the TX-side clock and data recovery (CDR) retimes the data to reduce jitter and produce a clean data output. The driver further increases the amplitude of the data signal to make an optimum input condition for the MZI. The MZI is based on a reverse-biased p-n junction for high-speed operation at the cost of reduced phase shift, thus reducing the extinction ratio (ER). ER can be raised by increasing the length of the modulator or enlarging the driver output swing. The driver output swing and the length of the modulator in this design are 5 V_{ppd} and 4 mm, respectively. A holographic lens is used as a fiber-optic coupler for both the TX

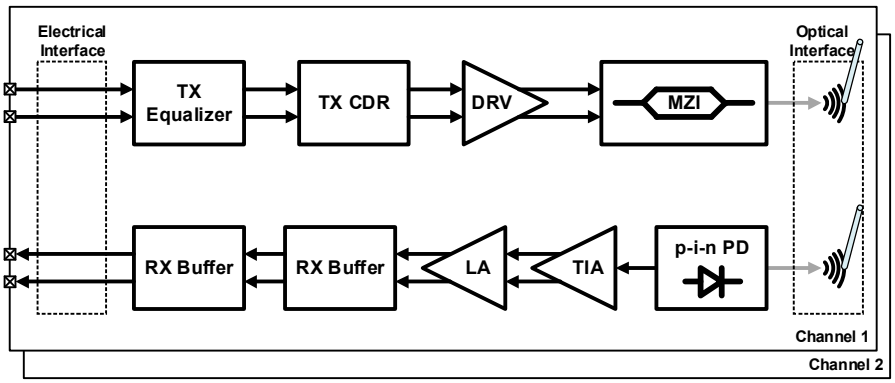


Fig. 2.3 2 x 10-Gb/s dual channel optoelectronic transceiver [29].

and the RX. The PD at the RX is flip-chip bonded to the silicon die. The transimpedance amplifier (TIA) converts the PD current to a voltage which is further amplified by a limiting amplifier (LA) for the proper operation of the CDR. Following the CDR, the output buffer is employed to drive a differential 100-Ω transmission line. In 2007, Luxtera further developed their photonic device library by adding a dense WDM (DWDM) building block. By implementing DWDM multiplexers and demultiplexers, they succeeded in integrating a 4 x 10-Gb/s DWDM optoelectronic transceiver in a 0.13-μm CMOS SOI technology [30].

The silicon photonics research group at MIT has also published successful works in monolithic optical transceivers. In 2012, the first monolithic optical receiver in a sub-100-nm standard SOI process was implemented [31]. In this work, novel approaches are proposed toward implementing a PD and a detecting scheme, which can be attributed to the increased design flexibility of the PD in the monolithic integration platform. An important advantage of the monolithic integration is the low wiring capacitance for the PD. An integrating receiver, described in Figure 2.4, can be the most appropriate topology by exploiting the low-capacitance condition and

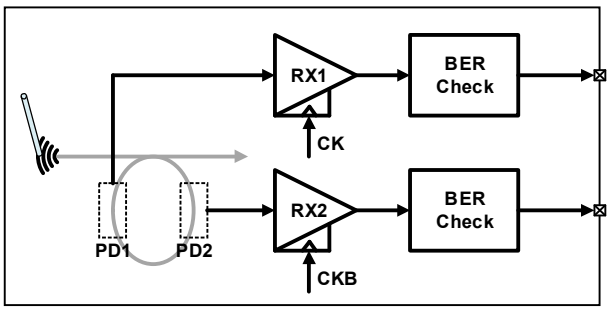


Fig. 2.4 Monolithic receiver with PD splitting [31].

the presence of an RX clock. However, an insufficient margin for evaluation can severely degrade the sensitivity. PD splitting combined with a double-data-rate (DDR) scheme is implemented in this work to alleviate the timing margin. By interdigitating the metal contacts of the PD, it can be divided into two separate PDs sharing a common waveguide, thus reducing the photocurrent of each path by half. In spite of the halved photocurrent, the PD splitting is more advantageous at higher data rates considering that the sensitivity degrades exponentially with the shortened evaluation time. Another improvement is in the PD structure that incorporates a ring resonator to confine the light, thereby enhancing the absorption with a shorter PD length. The implemented receiver chip operates at the data rate of up to 3.5 Gb/s with a good energy efficiency of 50 fJ/b. This demonstration is significant in that it paved the way for future memory-processor interfaces based on silicon photonics.

A monolithic optical transceiver operating at higher speed was also reported by UC San Diego and Oracle in the same year [32]. The chip is implemented in a 0.13-μm SOI CMOS technology and achieves a 25-Gb/s operation with an energy efficiency of 10.2 pJ/b for the entire transceiver. The optical TX is based on a micro-ring modulator and the optical RX employs a germanium PD. Notably, an asymmetric pre-emphasis is used even for the reverse-biased ring modulator. When the optical devices are forward-biased, or more specifically, when operating a VCSEL or a ring modulator in a forward-biased condition, the storage time of minority carriers degrades the operating speed and causes asymmetric rising and falling times. On the other hand, in the case of a reverse-biased ring modulator, the photon lifetime dominates the maximum operating speed, which also incurs asymmetry. Therefore, the asymmetric pre-emphasis, or independently controlling the rising and falling edges,

is essential even for the reverse-biased condition in order to improve signal quality at higher speed.

In 2015, a more sophisticated work was performed by collaboration among several research groups in academia and industry as shown in Figure 2.5 [33]. They demonstrate a good feasibility of future processor-memory links by implementing a processor, a memory, interface circuits, and photonic devices altogether onto a single microchip in a commercial 45-nm CMOS SOI process without any changes to the foundry process. The implemented chip reliably integrates 70 million transistors and 850 photonic components. The functionality of the processor-memory link is also verified by demonstrating actual read and write operations between the processor and the memory at 2.5 Gb/s. Although only a single-wavelength operation is demonstrated, the total aggregate bandwidth can be increased by more than 10 times without using additional fibers. A single external laser source is employed to supply

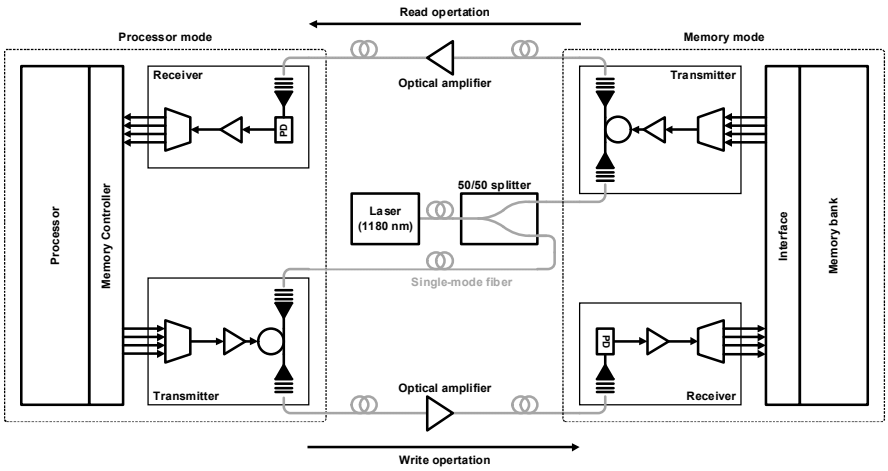


Fig. 2.5 Block diagram of memory-processor optical link [33].

1180-nm continuous wave (CW) for both the processor and the memory sides using the 50/50 power splitter. The receiver incorporating a SiGe PD exhibits a sensitivity of – 5 dBm for the bit-error rate (BER) of 10^{-12}. At the transmitter, a micro-ring resonator with reverse bias is employed and shows an ER of 6 dB. Since the resonant wavelength of the ring structure is sensitive to not only physical dimension but also temperature, a continuous stabilization of the wavelength is essential. In this work, a small part of the modulator output power is monitored and the resonant wavelength is tuned in a way that maximizes the output power by digitally controlling an embedded resistive heater inside the ring. Despite several limitations, the contribution of this work is evident and further development is expected to realize more practical silicon photonic systems in the near future.

On the other hand, the bulk CMOS platform has been relatively less popular than the SOI counterpart primarily due to the difficulty of implementing low-loss silicon waveguides. However, the realization of silicon photonics in bulk CMOS would be preferred because the bulk CMOS technology has been the mainstream so that most electronic parts are optimized in bulk CMOS processes with the lowest cost. On the other hand, the floating-body effect of SOI CMOS may lead to nonlinear characteristics of the active transistors and heat dissipation can be problematic, thus rendering

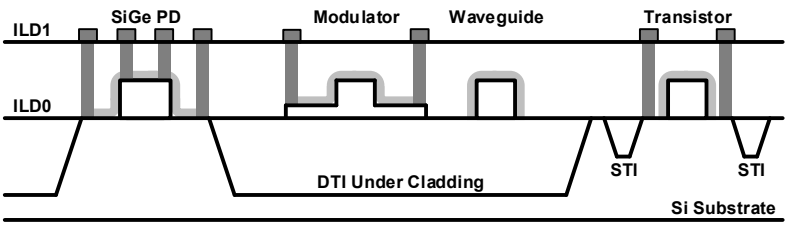


Fig. 2.6 Cross section of photonic and electronic devices [34].

the SOI platform inadequate for the extremely high level of integration. Therefore, photonic systems on the bulk CMOS platform have also attracted significant attention. In 2014, the first monolithic integration of photonic and electronic devices on the bulk CMOS was reported [34]. The most challenging part is the waveguide integration with low propagation loss. Figure 2.6 shows the cross section of the implemented silicon photonic devices. For the formation of a polysilicon-based waveguide, a lower cladding and an upper cladding are implemented using deep-trench isolation (DTI) and silicon nitride barrier films between the inter-level dielectric (ILD) layers, respectively. Several optimization processes such as the crystallization of amorphous polysilicon and low-pressure chemical vapor deposition (LPCVD) of silicon nitride are also performed to further reduce the loss of the waveguide. The resulting waveguide achieves a record attenuation of 10.5 dB/cm which is sufficiently low to support resonant ring modulators. A polysilicon resonant detector and a SiGe p-i-n detector are both implemented. The fully functional optical transceiver operates at 5 Gb/s exhibiting an energy efficiency of 2.8 pJ/b.

Subsequently, a more advanced work was demonstrated by incorporating a DWDM system into the bulk CMOS platform [35]. A DWDM optical transceiver is monolithically integrated on a 0.18-μm bulk CMOS process with all the optical devices implemented using polysilicon without relying on epitaxial crystallization of silicon or the introduction of germanium. By applying the same approaches presented in [34], low-loss photonic devices including waveguides, couplers, micro-ring modulators, and PDs can be successfully integrated with minimal modifications to the original CMOS process. Moreover, based on the micro-ring structures, 9-wavelength TX and RX DWDM macros are also realized as shown in Figure 2.7. The TX employs an internal pseudo-random binary sequence (PRBS) generator, a serializer, and a modulator driver, enabling electro-optic conversion by the carrier-depletion micro-ring modulator. At the RX side, a polysilicon-based PD utilizing absorption from defect states, combined with the resonant structure, provides an acceptable responsivity of 0.2 A/W. Similarly, the PD splitting technique in [31] is also employed to mitigate the timing margin. After the DDR-based receiver front-

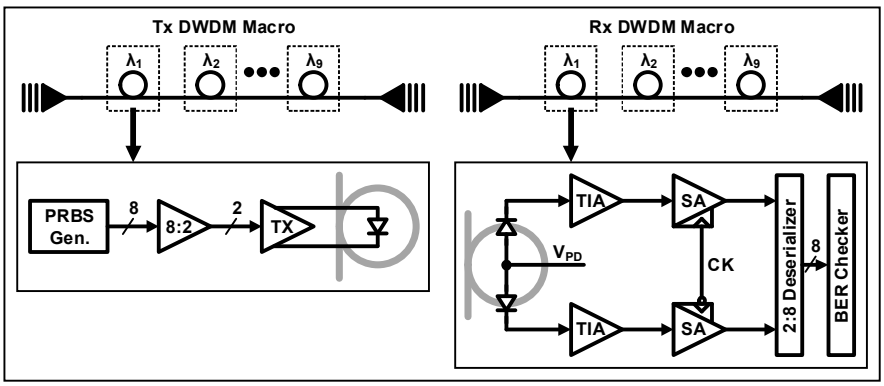


Fig. 2.7 Monolithically integrated DWDM optical link in bulk CMOS [35].

end, the data is further de-serialized for the estimation of BER. In order to achieve a stable wavelength-locking of the DWDM system, a heater and a digital logic controlling the heater are also implemented. Consequently, for a total aggregate data rate of 45 Gb/s, a 9-wavelength DWDM link sharing a single bus waveguide is successfully demonstrated.

Silicon photonics based on bulk-Si substrate is also an attractive option for processor-memory interfaces, especially for cost-sensitive dynamic random-access memory (DRAM) applications [36]. Processor-memory interfaces are currently facing two challenging industrial requirements. The first requirement is the demand for higher bandwidth, according to which the next-generation DRAM (DDR5) standard requires the maximum per-pin data rate of 6.4 Gb/s. The other requirement is that the total memory capacity that a memory controller can handle should be maximized. However, the existing schemes such as the multi-drop bus topology or the point-to-pint bus topology, cannot satisfy the aforementioned requirements simultaneously. Fortunately, if an optical interconnection is applied to the multi-drop bus interface, the per-pin data rate can be increased significantly without degrading the total available memory capacity. The CMOS SOI platform can never be adopted in the DRAM

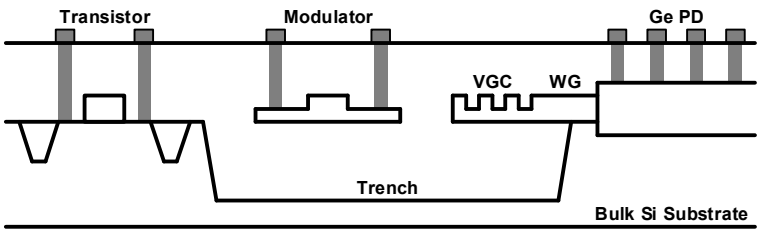


Fig. 2.8 Schematic view of EPIC structure [36].

interfaces where cost is a crucial factor. Samsung has presented several works based on the bulk-Si platform, which is different from the platform in [35] mainly in terms of waveguide-fabrication techniques. The schematic view of integrated EPICs is shown in Figure 2.8. Unlike in [35], the waveguides are formed through a local crystallization based on epitaxy. For that purpose, a trench is formed first and filled with silicon dioxide. Subsequently, amorphous silicon (a-Si) is deposited using LPCVD. After the deposition, a-Si is crystallized by solid-phase epitaxy (SPE) using the bulk-Si substrate near the edges of the trench as a crystal seed, thereby growing the crystal laterally toward the center of the trench. Finally, the waveguide is patterned using dry etching. Owing to the introduction of epitaxy, a higher quality of crystallization as compared with the previous works is obtained, exhibiting a waveguide loss of only 3 dB/cm. A further improvement can be achieved using a laser-induced epitaxial growth based on the liquid phase epitaxy method, resulting in an almost perfect crystallization, so that the grain size is comparable to that of the bulk Si. Using the waveguides formed by the SPE method, both the MZI and the micro-ring modulators are realized. Two types of the PD are also implemented by introducing Ge on Si: a surface-illumination type and a butt-coupled-waveguide type. Although the photonic devices alone exhibit good performance, overall performance degradation is observed when integrating the EIC and the PIC together with the combined process, which needs to be improved through process optimizations. Samsung has also demonstrated the feasibility of a multi-drop bus topology based on silicon photonics. Figure 2.9 describes an optical link with the hybrid integration of the EIC and the PIC, which shall be developed into a fully monolithic EPIC in the future. The implemented optical link consists of two sets of TXs and RXs processing 8 DQ signals,

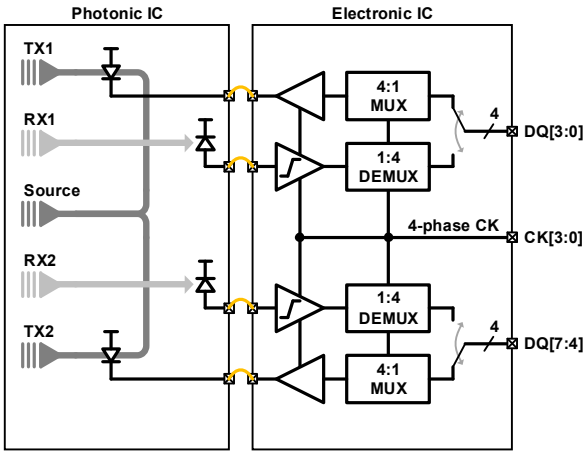


Fig. 2.9 Block diagram of optical transceiver [36].

and simply emulating a read or a write operation of the controller-memory communication. The link successfully demonstrates an error-free operation at the data rate of up to 2.5 Gb/s per fiber.

2.3 Hybrid Integration

As briefly discussed so far, silicon photonics has undergone remarkable progress, demonstrating its potential to replace the legacy copper-based interconnections. Nevertheless, some issues have yet to be addressed to prevent degradation of the EPIC performance. Moreover, the compatibility with a FinFET CMOS platform is unknown and should be investigated in the future. Alternatively, a hybrid integration leveraging the respective process optimizations of the EIC and the PIC would be

more advantageous at present, and it has become viable owing to advanced 3-D integration techniques. In this section, several recent works based on the hybrid integration will be reviewed and compared with the monolithic integration in various aspects.

The most common integration method is the use of a wire-bonding technique which has been prevalently used in EIC packages. However, wire bonding severely degrades signal integrity as the data rate becomes higher, because the length of a wire is directly translated into an inductance. For even higher data rates, the situation degrades as the wavelength of the signal becomes comparable to the length of the wire. Due to its relatively large dimension, wire bonding also limits the maximum pin density; hence it is not appropriate for modern system-on-chip packages where an extremely high pin density is required. Nevertheless, thanks to its distinguishably low-cost characteristic, it is still widely used for low-speed packages and chip-on-board test environments. An example of the integration of EPIC based on wire bonding is shown in Figure 2.10. A typical PD can be wire-bonded to a TIA chip and the output of the TIA is connected to a trace on a printed circuit board (PCB) to interface with end-launch connectors. Based on wire bonding, the works in [37] and [38] achieve operating data rates of 25 Gb/s and 64 Gb/s, respectively.

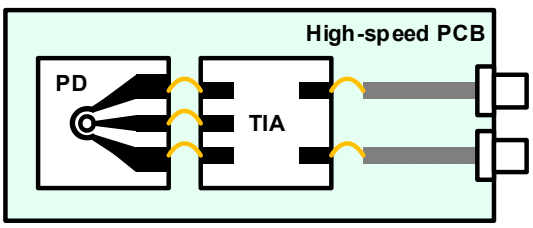


Fig. 2.10 Illustration of wire-bonded EIC and PIC.

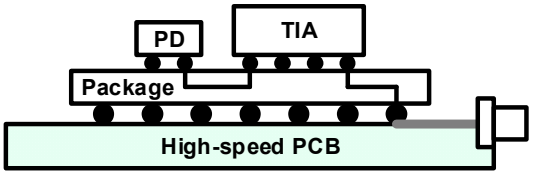


Fig. 2.11 Illustration of wire-bonded EIC and PIC.

Flip-chip bonding is a more advanced technique based on face-to-face bonding, completely eliminating the bond wires. Because of a shortened chip-to-chip space (several tens of μm with modern techniques), the parasitic effects can be significantly reduced, resulting in improved signal integrity at higher speeds. Figure 2.11 illustrates a typical flip-chip package. The PD chip and the TIA chip are flip-chip bonded on the same package substrate through which high-speed signal interconnection is made. The output of the TIA is connected to the PCB trace via the package and the solder bump. In this case, the structure of PICs should be a back-illumination type for proper fiber coupling. With this configuration, the work in [39] presents a 4 x 28-Gb/s optical receiver exhibiting a good sensitivity.

Similarly, but in a slightly different manner, the EIC and the PIC are stacked vertically using an interposer as shown in Figure 2.12, [40], realizing a 12 x 5 two-dimensional EPIC array in order to maximize the I/O density. The implemented transceiver chip achieves a total aggregate data rate of as high as 600 Gb/s.

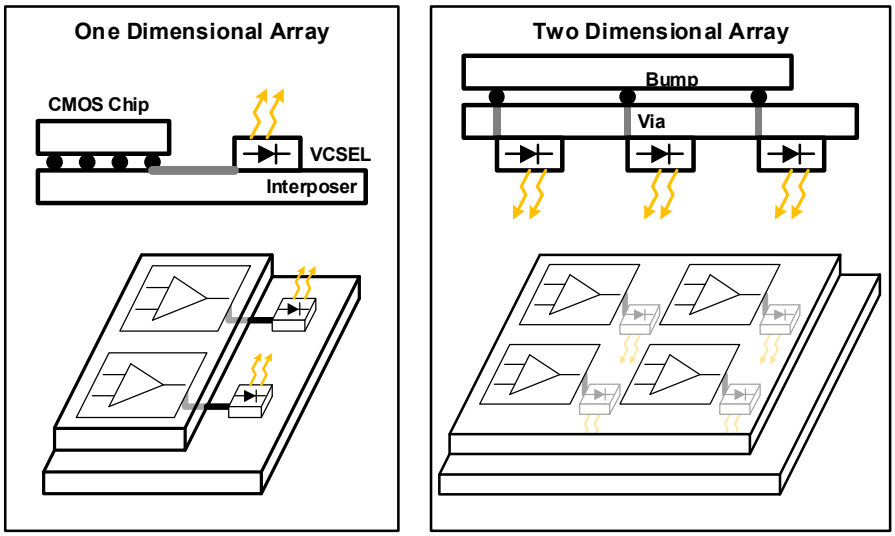


Fig. 2.12 Implementation of 2-D array of EPIC [40].

Figure 2.13 describes an improved flip-chip package that directly bonds the EIC and the PIC using micro bumps, thus obviating any redistribution layer between high-speed regions to further reduce the parasitic effects [41]. The rest of the signals are connected to the package substrate through C4 bumps. An additional cost may be incurred by such a packaging technique due to the requirement of a customized substrate.

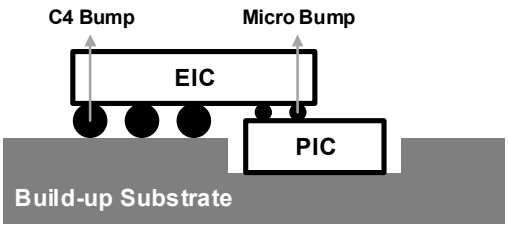


Fig. 2.13 Flip-chip bonded EIC directly on PIC using micro bump [41].

Currently, the most advanced technique is shown in Figure 2.14. Contrary to the previous techniques, it uses the PIC as a redistribution layer itself, thus simplifying the packaging process [25], [28], [42]. The EIC is flip-chip bonded to the macro PIC in a similar way. Not only the high-speed signal pads but also the low-speed signal pads are connected to the PIC. Subsequently, the low-speed signals are carried outside the PIC through a typical wire-bonding package. Based on this technique, the work in [28] demonstrates a 4 x 20 Gb/s WDM transceiver and [25] implements a 56-Gb/s optical transmitter with the MZI structure.

In Table 2.1, we summarize and compare the optical transceivers presented so far. Notably, the overall performance of the hybrid-integrated transceivers is better than that of the monolithic transceivers, especially in terms of energy efficiency. This can be attributed to the independently optimized processes for the EIC and the PIC. However, it should also be noted that recent monolithic transceivers have demonstrated significant improvement, and are comparable to their hybrid counterparts.

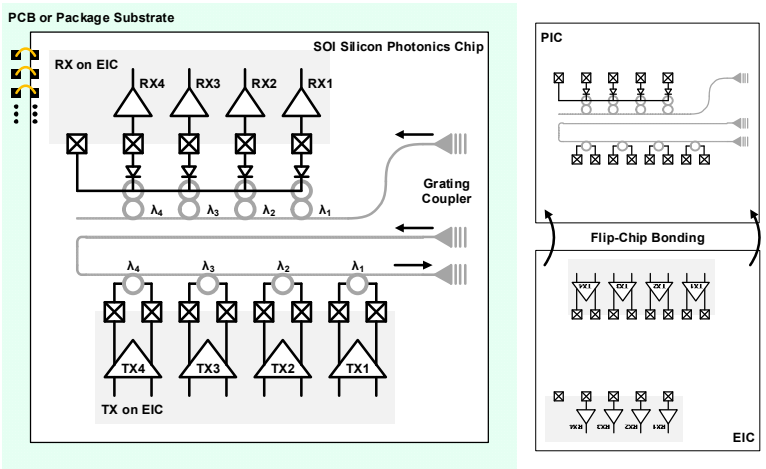


Fig. 2.14 EIC flip-chip bonded to macro PIC [28].

Table 2.1 Summary of monolithically- and hybrid-integrated optical transceivers.

	[29]	[30]	[32]	[35]	[37]	[40]	[41]	[28]
Integration method	Monolithic	Monolithic	Monolithic	Monolithic	Hybrid	Hybrid	Hybrid	Hybrid
Technology (EIC)	130nm SOI	130nm SOI	130nm SOI	180nm Bulk	90nm Bulk	65nm Bulk	28nm Bulk	40nm Bulk
Technology (PIC)					GaAs	GaAs	28nm SOI	130nm SOI
Wavelength	1535-1555nm	1549-1554nm	1560nm	1280-1295nm	850nm	850nm	1556nm	1550nm
# of WDM channels	N/A	4	N/A	9	N/A	N/A	N/A	4
Max. data rate/CH	10Gb/s	10Gb/s	25Gb/s	5Gb/s	25Gb/s	10Gb/s	25Gb/s	20Gb/s
Total throughput	20Gb/s	40Gb/s	25Gb/s	45Gb/s	25Gb/s	600Gb/s	25Gb/s	80Gb/s
TX								
Modulation type	MZI	MZI	MRR	MRR	VCSEL	VCSEL	MRR	MRR
ER	5-6dB	>4dB	6.9dB	6.9dB	5.1dB	5.6dB	6.5dB	> 7dB
Power/CH	N/A	575mW	208mW	N/A	46mW	69.5mW	72.5mW	32.3mW
Energy efficiency/CH	N/A	57.5pJ/b	8.32pJ/b	N/A	1.84pJ/b	6.95pJ/b	2.9dB	1.6pJ/b
RX								
PD type	N/A (external)	N/A (external)	Ge waveguide	Si (defect) waveguide	GaAs Top-illumination	N/A	N/A	Ge waveguide
Sensitivity (BER of 10^{-12}) @ max. data rate	-19.5dBm	-15dBm	-6dBm	-7.5dBm	-6dBm @ 22Gb/s	-16dBm (estimated)	-8dBm	-7.2dBm
Power/CH	N/A	120mW	48mW	N/A	44.4mW	68.2mW	50mW	11.6mW
Energy efficiency/CH	N/A	12pJ/b	1.92pJ/b	N/A	1.78pJ/b	6.82pJ/b	2pJ/b	0.73pJ/b
Total power	2.5W	3.5W	256mW	675mW	90.4mW	8.26W	122.5mW	175.6mW
Total energy efficiency	125pJ/b	87.5pJ/b	10.2pJ/b	15pJ/b	3.62pJ/b	13.77pJ/b	4.9pJ/b	2.2pJ/b

2.4 Silicon-Based Photodiodes

In this section, we examine the basic terminologies of a PD. We also briefly discuss several PD structures based on silicon photonics and their operating principles.

2.4.1 Basic Terminology

We first present some terminologies related to the PD characteristics. The quantum efficiency is the ratio of the number of generated electro-hole pairs to the number of incident photons. Since the quantum efficiency directly determines the sensitivity of the entire receiver, designing a PD to have a high quantum efficiency is crucial. This quality is generally expressed as

$$\eta = \frac{I_{ph}/q}{P/hv}. \tag{2.1}$$

The internal quantum efficiency is sometimes defined by de-embedding the loss occurring at the detector interface to evaluate the detector performance alone. The responsivity is a more frequently used quantity by engineers with a similar meaning to the quantum efficiency, and can be expressed as

$$R = \frac{I_{ph}}{P} = \frac{\eta q}{hv} = \eta \frac{\lambda(nm)}{1240}. \tag{2.2}$$

The bandwidth of the PD is determined by two time constants: the transit time and the RC time constants which arise from the series resistance and the junction capacitance of the PD. Hence, the total time constant should be considered for the estima-

tion of the bandwidth as follows.

$$\tau = \sqrt{\tau_{transit-time}^2 + \tau_{RC}^2} \ . \tag{2.3}$$

2.4.2 Silicon PD

Silicon is a good waveguide material for telecommunication wavelengths due to its large bandgap energy equivalent to the energy of light with the wavelength of approximately 1100 nm. Therefore, ironically, silicon cannot be a good absorber for wavelengths longer than 1100 nm. Fortunately, some promising results have been obtained by enabling a sub-bandgap detection without introducing any other materials. It is known that silicon can also detect light of wavelength longer than 1100 nm if crystal defects are present, which cannot be easily explained theoretically, but can be proven experimentally [43]–[47]. By deliberately introducing defects into a p-n diode in a waveguide structure, the sub-bandgap detection in silicon can be realized. In [43], a silicon waveguide PD is implemented, wherein the quantum efficiency is enhanced by applying ion implantation. The implemented waveguide PD is described in Figure 2.15. The diode is formed on a SOI substrate using a standard CMOS process. After the formation of the waveguide PD, the ion implantation is conducted to introduce the crystal defects. The lengths of the waveguides are 0.25 and 3 mm. The PD with a waveguide length of 0.25 mm shows a responsivity of sub-0.1 A/W at low reverse voltages; however, the responsivity can approach 1 A/W by increasing the reverse voltage at the cost of increased dark current, thereby degrading the minimum detectable power (MDP). The PD with a waveguide length of 3 mm can absorb virtually all the light, thus enhancing the responsivity at lower re-

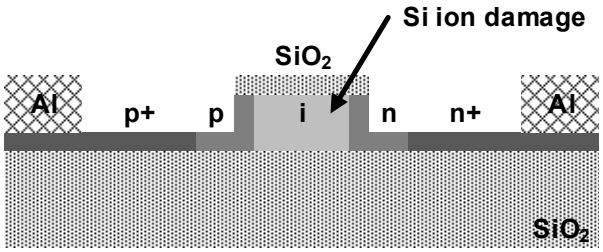


Fig. 2.15 Cross-sectional view of waveguide PD [43].

verse voltage. However, the increased length directly results in higher junction capacitances, which severely degrades the frequency response. The estimated bandwidth of the 0.25- and 3-mm PDs are 10-20 GHz and 2 GHz, respectively. Further research improves the performance of the silicon waveguide PD through several process optimizations, providing a bandwidth of > 35 GHz and an internal quantum efficiency of 0.5 to 10 A/W [44].

As clearly observed in [43] and [44], the silicon waveguide PD suffers from the tradeoff between the quantum efficiency and the bandwidth. In order to overcome this severe tradeoff, the work in [45] suggests that the adoption of a ring resonator can enhance the absorption with significantly reduced device dimension. As shown in Figure 2.16, the resonator-enhanced PD achieves a responsivity of 0.14 A/W with a length 10 times shorter than that of the straight waveguide for the same responsivity. The dark current level is also maintained below 0.2 nA, thus exhibiting an MDP of only 1.4 nW. Moreover, a resistive heater is employed for tuning the resonance wavelength.

The silicon PDs discussed so far are all based on the SOI platform. However, the

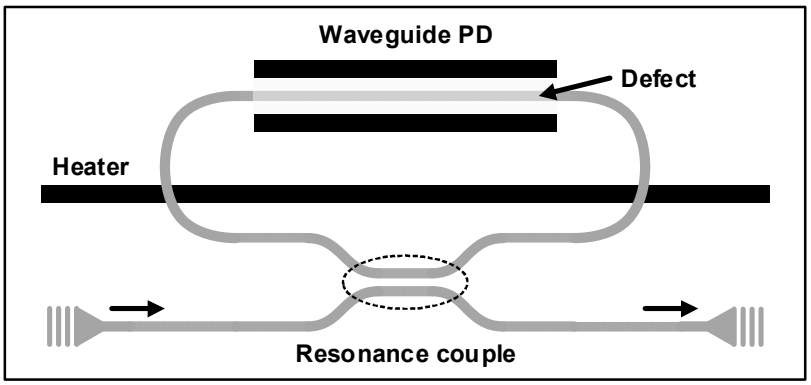


Fig. 2.16 Schematic view of silicon resonator-enhanced PD [45].

realization of the silicon PD on the bulk CMOS platform would be essential for a truly monolithic silicon photonic receiver. In [46], starting with a bulk silicon substrate, a micro-ring resonator PD using polysilicon is demonstrated, achieving a responsivity of 0.15 A/W, a dark current of 40 nA, and a gigahertz frequency response. Recently, the work in [47] achieves further improvements of overall performance by applying mid-level implants to a typical p-i-n structure. The recent advances in silicon PDs are significant because they simplify the overall process with a minimal addition of steps, indicating a better compatibility with the standard CMOS processes than their Ge-based counterparts. Especially, the silicon PD incorporated in a waveguide has a distinct advantage that the integration with a waveguide is readily realized, which is challenging for the Ge-based detectors. Furthermore, the resonator-based structure directly enables the WDM function which is a desirable feature for highly-dense interconnection.

2.4.3 Germanium PD

Another approach involves the introduction of Ge which has a smaller bandgap energy than Si. By appropriately combining Si with Ge, the detectable wavelength can be extended from 1100 nm, with Si alone, to longer than 1550 nm, covering all the telecommunication wavelengths. Furthermore, Ge exhibits higher mobility of electrons and holes, resulting in a faster detection. Therefore, Ge has long been considered as an ideal candidate to replace the conventional III-V detectors. However, the adoption of Ge is severely restricted due to the mismatch of the lattice constants between Si and Ge. The lattice mismatch poses a constraint on the maximum Ge thickness that can be grown on Si without introducing crystal defects [48]. Consequently, Ge thickness is limited for maintaining the dark-current level as low as possible, resulting in a low quantum efficiency. Hence, the epitaxial growth of a thick Ge layer on Si should be the key to the realization of the SiGe PD and, more importantly, it should be achieved in a CMOS-friendly manner.

An intuitive way to obtain a thick Ge layer with minimal dislocations is by using a graded buffer layer [49]. Based on this technique, the work in [49] achieves an extremely low dark-current level which is comparable to the theoretical reverse saturation current. Even though it completely addresses the problem of lattice mismatch, it usually results in a tall PD structure; consequently, this technique is not compatible with CMOS back-end processes and waveguide coupling. Alternatively, a direct growth of Ge on Si with the aid of a thin Ge buffer layer can be more effective [50]. Starting from a thin Ge buffer layer at a low temperature of 350°C, the Ge layer can be grown to be sufficiently thick for the absorption of infrared light at a relatively

high temperature of 600°C, thus avoiding the problem of crystal defects. In this way, the detection of 1300-nm wavelength can be successfully achieved with a responsivity of 240 mA/W. Although not reported in this work, the dark current level is estimated to be high because the defects are not completely eliminated. In [51], it is demonstrated that the dislocation density can be remarkably reduced by applying cyclic thermal annealing at a high temperature of 900°C. Further extending the work in [50] by employing a cyclic thermal annealing process, the work in [52] achieves a high responsivity with a low-dark current level and the frequency response is suitable for gigabit operation. However, high-temperature annealing may not be compatible with the standard CMOS process, necessitating a different defect-handling technique for the seamless integration of the EPIC. The work in [53] demonstrates that the selective growth of Ge through multiple hydrogen annealing is also a good approach to reduce the dislocation density.

2.4.4 Integration with Waveguide

PDs can be categorized into two types according to coupling schemes: a free-space coupled PD and a waveguide-coupled PD as described in Figure 2.17. In the case of the free-space coupled PD, a fiber can be directly coupled for a top illumination or a back illumination. In either case, the direction of the incoming light is always parallel to that of the carrier collection. Therefore, in order to enhance the quantum efficiency, the thickness of an absorption layer should be sufficiently large, which results in an increased transit time, and hence degrades the frequency response. On the other hand, in the waveguide-coupled PD, the tradeoff between the

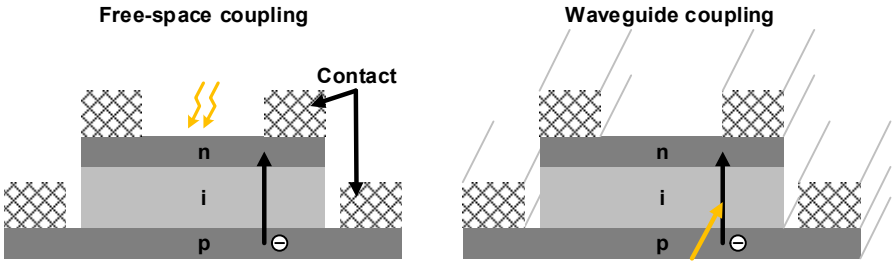


Fig. 2.17 Categorization of PD by coupling schemes.

quantum efficiency and the bandwidth can be relaxed. This is because the absorption can be improved by increasing the absorption length along the direction of the light, which does not affect the transit time. However, in this case, the increased junction capacitance may degrade the bandwidth, but not as much as in the case of the free-space coupled PD. Another benefit is that the waveguide PD is more suitable for an on-chip WDM system, thus obviating the use of additional fibers. Evanescent coupling and butt coupling are most commonly used for waveguide coupling. Evanescent coupling is easy to realize whereas it is relatively difficult to design a structure for butt coupling. However, in terms of quantum efficiency, butt coupling with a well-designed waveguide is generally much better than evanescent coupling.

Chapter 3

Circuit Techniques for Optical Receiver

3.1 Basis of Transimpedance Amplifier

The TIA is the first electronic circuit that directly interfaces with a PD for current-to-voltage conversion. Therefore, the overall performance of the receiver is predominantly determined by that of the TIA. There are four main parameters defining the performance of the TIA: gain, bandwidth, noise, and power. Further, these performance parameters are strongly interdependent and trade-offs exist among them. For example, it is not easy to improve the gain and the bandwidth simultaneously without sacrificing the power. Thus, a proper compromise should be made when determining gain, bandwidth, and power consumption so that the receiver

achieves a targeted sensitivity. Before discussing the TIA circuits, we look over some basic terminologies essential to understanding the optical receiver. Let us assume that a simple optical receiver is configured as shown in Figure 3.1. The PD receives an optically signal that is characterized by an optical power and an extinction ratio. The optical power can be defined as either an average power or an optical modulation amplitude (OMA) which is the difference between P_1 and P_0. The extinction ratio (ER) is a ratio of P_1 to P_0, which is solely determined by the transmitter. Thus, the OMA is a useful metric when evaluating the receiver performance alone with the transmitter performance de-embedded. The optical power is then converted to a current in a ratio that is determined by the PD, also called a responsivity. Subsequently, the photocurrent is amplified and converted to a voltage by the TIA. The

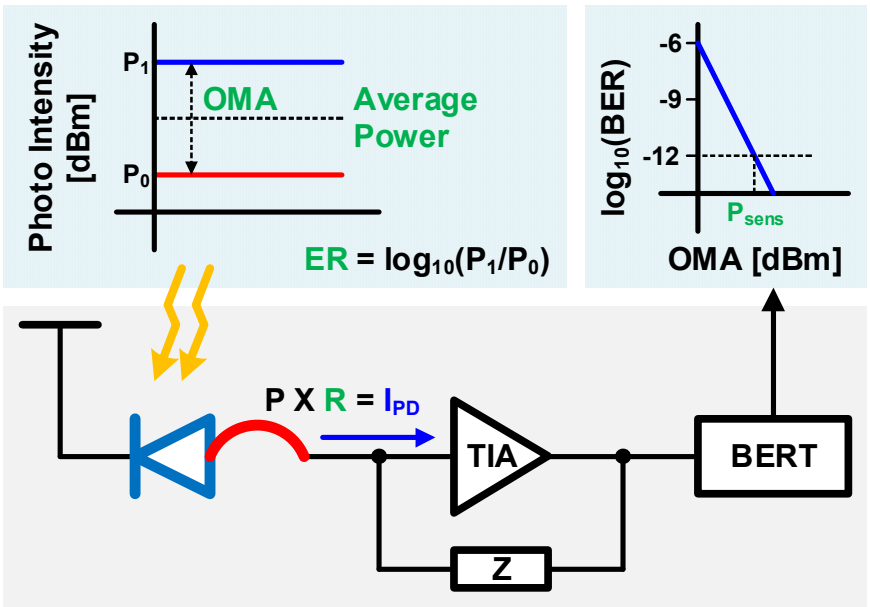


Fig. 3.1 Basic terminologies for optical receiver.

output of the TIA is fed to a bit-error-rate tester (BERT) for an error detection. The sensitivity is defined as the minimum required input optical power for the BER of less than a certain value (typically 10^{-12}). Because the sensitivity is the most important performance metric, it is also crucial to understand the relation between the sensitivity and the noise performance of the TIA. The noise sources surrounding the TIA are illustrated in Figure 3.2. The TIA has to accommodate not only the PD noise but also the circuit noise and all the noises can be referred to the input of the TIA for an SNR calculation. Because the PD shot noise power is proportional to the optical signal power, the total noise power is data-dependent. The expressions of the total noise power for both the data of '1' and '0' can be given as follows:

$$\sigma_1^2 = 2e \cdot BW \cdot I_1 + \overline{I_{n,TIA}^2} \,. \tag{3.1}$$

$$\sigma_0^2 = 2e \cdot BW \cdot I_0 + \overline{I_{n,TIA}^2} \,. \tag{3.2}$$

With the noise taken into account, the probability of receiving '1' or '0' can be calculated as shown in Figure 3.3(a). From this result, one can estimate the BER as

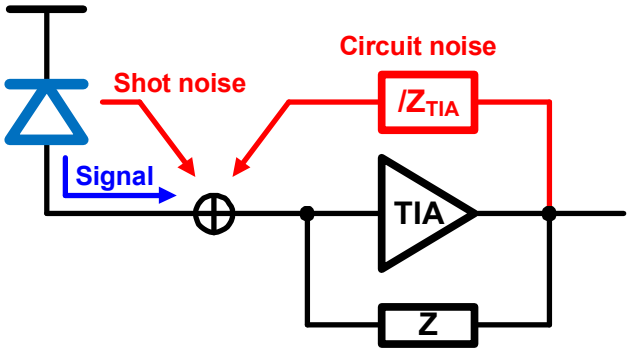
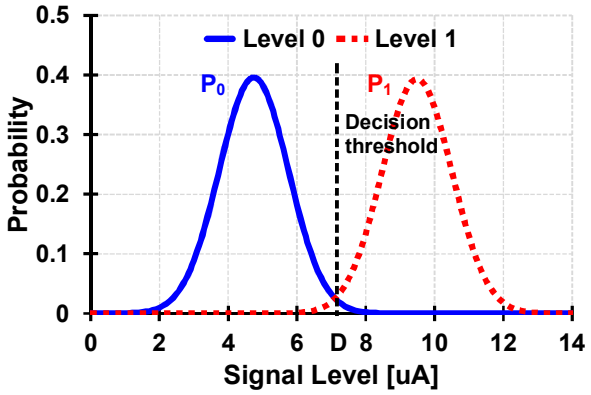


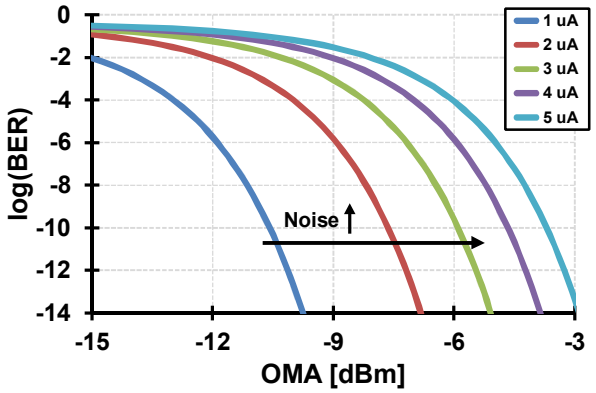
Fig. 3.2 Noise sources of TIA.

$$BER = \frac{1}{2}\left[\int_{-\infty}^{D} p_1(x)dx + \int_{D}^{\infty} p_0(x)dx\right] = \frac{1}{2}erfc\left[\frac{I_1 - I_0}{\sqrt{2}(\sigma_1 + \sigma_0)}\right] \quad (3.3)$$

The sensitivity can be estimated with various TIA noise levels as shown in Figure 3.3(b). Such an estimation is a prerequisite for an optimum design of the TIA.



(a)



(b)

Fig. 3.3 (a) Probability of receiving '1' or '0' and (b) sensitivity estimation.

3.2 Topology of TIA

Traditionally, compound semiconductor devices were preferred for implementing optical interface circuits due to their high-speed capability. However, significant improvements in CMOS technologies have reduced the performance gap between these two technologies. Moreover, recently, further advancements in circuit technologies have led to CMOS implementations of extremely high-speed optical receivers whose operating speeds are higher than 40 Gb/s [38], [54].

Besides enjoying the advantage of CMOS scalability, the success of "CMOS-compatible" silicon photonics also mandates the CMOS implementation of electronic circuits. Although most of optical receivers still rely on hybrid integration which interconnects multiple ICs from different technologies through bonding, monolithic integration is gaining attention and some remarkable results have been demonstrated as previously mentioned. If realized, monolithic integration would be the best solution since it provides many benefits, such as enhanced signal integrity, small area, and reduced packaging cost, when compared to hybrid integration.

In this section, we focus on CMOS realizations of the TIA by providing a brief overview of various TIA topologies reported so far.

3.2.1 Resistor-based TIA

The most intuitive way to implement the TIA is employing a single resistor as shown in Figure 3.4. In this configuration, the calculations of the gain, R_T, and the

bandwidth, f_{-3db}, are straightforward and can be expressed as follows:

$$R_T = R_{TIA} . \tag{3.4}$$

$$f_{-3dB} = \frac{1}{2\pi R_{TIA} C_{PD}} . \tag{3.5}$$

As evident from the above equations, there is a severe tradeoff between the gain and the bandwidth. On the other hand, the total integrated noise at the output is given by

$$\overline{V_{n,out}^2} = \frac{kT}{C_{PD}} , \tag{3.6}$$

where k is the Boltzmann constant and T is the absolute temperature. Furthermore, the signal-to-noise ratio (SNR) can be defined to assess the performance of the TIA as follows.

$$SNR = \frac{C_{PD}}{kT} I_{in}^2 R_{TIA}^2 , \tag{3.7}$$

which implies that increasing R_{TIA} enhances the SNR indefinitely. However, a large R_{TIA} introduces an inter-symbolic interference (ISI), which limits the maximum value of R_{TIA} for a given C_{PD}. In general, the minimum bandwidth sufficiently suppress-

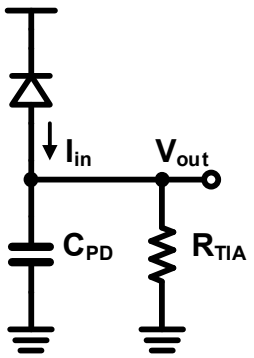


Fig. 3.4 Simple implementation of TIA using single resistor.

ing the ISI is chosen to be 0.5–0.7 times the target data rate. Thus, the only way to enhance the SNR is increasing the input optical power, thereby providing a large input current. Furthermore, a large C_{PD} renders this type of TIAs unsuitable for high-speed optical receivers. Nevertheless, this type of TIAs exhibits a distinct advantage in that it consumes virtually zero power. Recently, by effectively cancelling the ISI, a resistor-based TIA is successfully demonstrated at moderately high speed with low-power consumption [55].

3.2.2 Common-gate-based TIA

As discussed in the previous section, the resistor-based TIA suffers badly from the gain-bandwidth tradeoff, which leads to the limited achievable SNR. In order to overcome the disadvantage, a common-gate (CG) topology has been widely used.

The basic CG amplifier as a TIA is described in Figure 3.5. Assuming a dominant pole is located at the input, the gain and the bandwidth are expressed as follows:

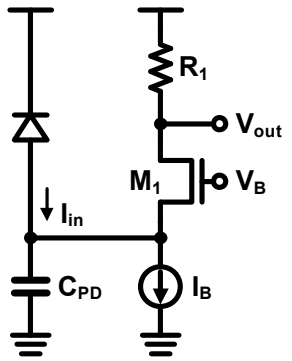


Fig. 3.5 Common-gate TIA.

$$R_T = R_1 . \tag{3.8}$$

$$f_{-3dB} = \frac{g_{m1}}{2\pi C_{PD}} . \tag{3.9}$$

Note that the gain-bandwidth tradeoff is now completely eliminated, indicating that the gain can be improved without degrading the bandwidth significantly.

Even if the SNR can be enhanced by employing the CG TIA, a direct tradeoff between the bandwidth and the power consumption exists, which limits the use of the CG TIA in the present configuration. Alternatively, a regulated-cascode (RGC) TIA shown in Figure 3.6 can further improve the bandwidth by lowering the input resistance [56]. The gain of the RGC TIA is the same as that of the CG TIA and the bandwidth is given by (with the same dominant-pole condition),

$$f_{-3dB} = \frac{g_{m1}\left(1+g_{m2}R_2\right)}{2\pi C_{PD}} . \tag{3.10}$$

With linearly increasing I_B, g_m and thus f_{-3dB} are increased approximately by the square of g_m, thereby alleviating the direct tradeoff between the bandwidth and the power consumption. Owing to their clear advantage, many TIAs based on the RGC

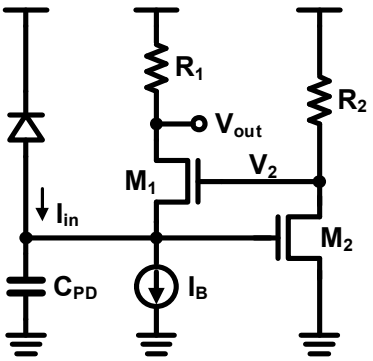


Fig. 3.6 Regulated-cascode TIA.

topology have been proposed. The bandwidth of the RGC TIA can be further enhanced by combining a shunt-shunt feedback [57] and a differential RGC TIA can also be configured [58]. Recently, RGC-based TIAs have been successfully demonstrated with operating speed higher than 25 Gb/s [41], [59], [60].

Despite the advantage of the RGC TIA, it is not suitable in the recent CMOS trend wherein the supply voltage is aggressively scaled down, because it suffers from a small voltage headroom due to stacking of the two NMOS transistors. Specifically, the output voltage, V_{out}, in Figure 3.6 has to accommodate the gate-source voltage of M_2, the drain-source voltage of M_1, and the voltage across R_1. The output of the common-source amplifier, V_2, also has to accommodate the gate-source voltages of M_1 and M_2, and the voltage across R_2. These conditions render the RGC TIA inapplicable to sub-1 V CMOS technologies.

As an alternative, a CG-feedforward TIA described in Figure 3.7 is proposed in [61], achieving a wider bandwidth and relaxing the voltage headroom simultaneously. Furthermore, the gain of the CG feedforward TIA is approximately the same as

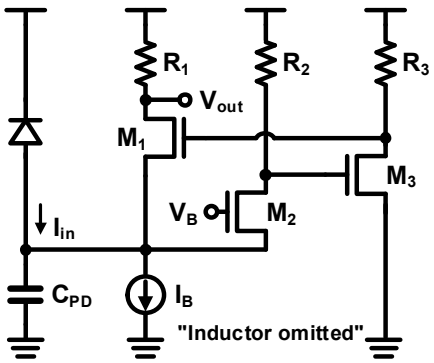


Fig. 3.7 CG-feedforward TIA.

those of the CG and the RGC TIA. The bandwidth can be calculated as,

$$f_{-3dB} = \frac{g_{m1}\left(1 + g_{m2}g_{m3}R_2R_3\right)}{2\pi C_{PD}}. \tag{3.11}$$

The bandwidth is significantly enhanced as compared to the RGC TIA and the voltage-headroom problems are also greatly mitigated. By employing this topology, the work in [61] achieved a considerably high bandwidth of 20 GHz in the standard CMOS platform.

3.2.3 Feedback-based TIA

Along with the CG-based TIAs, a shunt-shunt-feedback TIA has been widely employed in optical receivers. Figure 3.8 shows a basic feedback-based TIA topology. With an ideal amplifier, the gain and the bandwidth can be approximated as follows:

$$R_T = R_F. \tag{3.12}$$

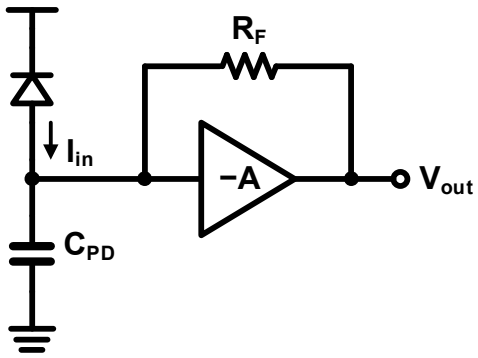
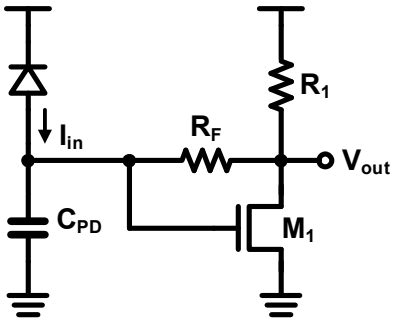


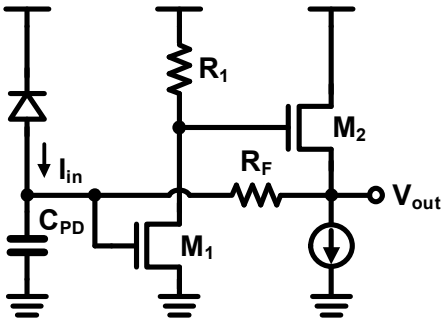
Fig. 3.8 Shunt-shunt-feedback TIA.

$$f_{-3dB} = \frac{1+A}{2\pi R_F C_{PD}} . \tag{3.13}$$

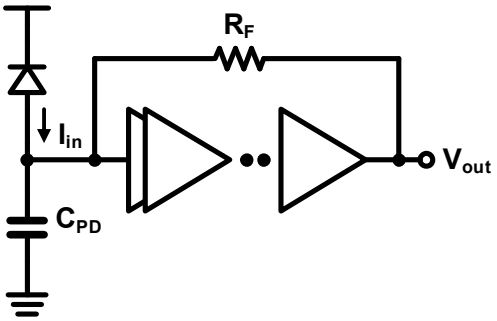
As (3.13) indicates, the bandwidth can be enhanced by the higher gain of the amplifier. Therefore, designing an amplifier that exhibits a high gain and a low output impedance is crucial for achieving good performance. In general, this type of TIAs is more advantageous than the CG-based topology in terms of voltage headroom and features a relatively simple architecture favorable for optimization. As shown in Figure 3.9, a variety of versions of the feedback-based TIA have been attempted with different implementations of the amplifier. Figure 3.9(a) describes a traditional TIA implementation which was first demonstrated with CMOS technology in [62] and achieved a 1-Gb/s operation. This topology has been widely used and has succeeded in achieving high-speed operations of 10 and 25 Gb/s [29], [63]. Alternatively, as shown in Figure 3.9(b), the combination of a common-source amplifier, a source follower, and a feedback resistor has been commonly employed to provide a low output impedance [64]–[66]. Especially, the work in [66] achieves a 10-Gb/s operation by combining this topology with bandwidth-enhancement techniques. In order to achieve a higher gain, the feedback can be applied to a multi-stage amplifier as described in Figure 3.9(c). In [67], the feedback resistor is connected between the input and the output of the three-stage inverter amplifier. The feedback can also be applied to a differential topology as implemented in [68] and [69] with a two-stage and a four-stage amplifiers, respectively. In spite of several advantages of the feedback-based TIAs, they suffer from an inherent stability issue which is particularly evident in the multi-stage topologies.



(a)



(b)



(c)

Fig. 3.9 Various implementations of shunt-shunt-feedback TIA.

3.2.4 Inverter-based TIA

Based on the shunt-shunt feedback, an inverter-based TIA is currently the most popular topology and it was first implemented with a CMOS technology for a 1-Gb/s operation [70]. As shown in Figure 3.10, the inverter-based TIA features a very simple architecture consisting only of a CMOS inverter and a feedback resistor. Even if the bandwidth enhancement factor is limited by the gain of the one-stage inverter, this topology has numerous advantages over its competitors. First, owing to its simple architecture, the voltage headroom is significantly mitigated and no biasing circuit is needed. Second, since it can exploit g_m of both the NMOS and the PMOS, high g_m can be achieved with considerably low power consumption [37]. Nevertheless, the inverter-based TIA has not been frequently used, because it is not compatible with the traditional technologies. However, there is a major contributor to the widespread use of this topology. The CMOS technology continues to be developed in such a way that most of the performance is optimized for digital circuits,

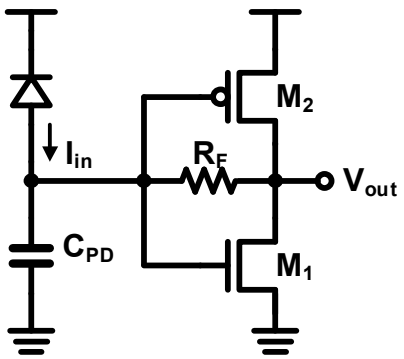


Fig. 3.10 Inverter-based TIA.

but not for analog counterparts, thereby forcing the inverter-based TIA to be the most suitable topology for modern CMOS technologies. Therefore, after being revisited in [37], the inverter-based TIA is now widely employed for high-speed and low-power applications [6], [28], [35], [38], [40], [54], [71]–[77]. In the state-of-the-art implementation, the optical receiver based on this topology demonstrates a 64-Gb/s operation, which is recorded as the fastest speed ever achieved in the CMOS platform [38].

3.2.5 Integrating Receiver

The aforementioned topologies are based on an amplifier, which offers a high-bandwidth and low-noise characteristic generally at the cost of an increased power consumption. In order to overcome this, an integrating receiver is proposed using a different method for achieving a high sensitivity with a low-power consumption [78]. The basic concept is briefly explained in Figure 3.11. The photocurrent is first integrated by a sampler using two non-overlapping phases, thus charging or discharging V_{in} according to the incoming data. After the integrating phase, the current value, V_n is compared with the previous value, V_{n-1}, to make a proper decision. Based on this scheme, the work in [78] achieves a 1.6-Gb/s operation while consuming a considerably low power. The works in [79] and [80] further extend this work, and achieve much higher operation speeds of 16 and 24 Gb/s, respectively.

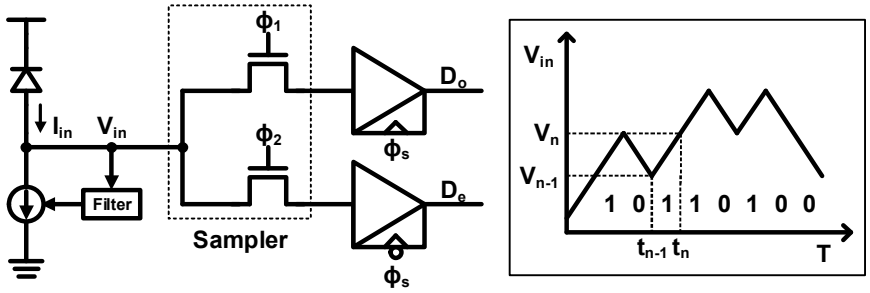


Fig. 3.11 Integrating receiver based on double sampling.

3.3 Bandwidth Extension Techniques

As previously discussed, the optical link provides a much higher bandwidth than the electrical link owing to its superior channel characteristic that the frequency-dependent loss is negligible. Actually, the speed limit does not come from the channel but from the circuit. Therefore, in order to take full advantage of the optical link, the speed limit of the circuit should be relaxed first. Fortunately, techniques for bandwidth extension have been massively investigated, enabling an extremely high-speed operation in the standard CMOS platform. In this section, a brief summary of various bandwidth extension techniques is presented.

3.3.1 Inductor-Based Technique

In this section, basic principles of inductive peaking techniques are described. An inductive peaking technique has a long history and its integration in CMOS technology is well documented in the literature [81]–[83]. Nevertheless, an integrated inductor occupies a huge silicon area so that it increases the cost of IC significantly. This is the reason why using inductors in optical receiver ICs should be considered carefully and the inductors should be optimized precisely. Fortunately, there are a lot of novel inductive peaking techniques, which have been developed to overcome the bandwidth limitation of the legacy electrical link or ultra-wideband (UWB) amplifier. Some of them have been adopted in the high-speed optical receiver during the last two decades. Basically, there are two types of inductive peaking, shunt peaking and series peaking, and their examples with a simple CS stage are shown in Figure 3.12(a). Without inductive peaking, the transfer function of the CS stage is

$$H(s) = \frac{g_m R}{1 + sRC}, \tag{3.14}$$

which is a simple one-pole system. On the other hand, for the shunt-inductive peaking which is shown in Figure 3.12(b), the inductor introduces a zero as well as an additional pole in the transfer function so that the transfer function becomes

$$H(s) = \frac{g_m (R + sL)}{1 + sRC + s^2 LC}. \tag{3.15}$$

By placing the zero properly in accordance with the positions of the poles, the zero can compensate the dominant pole, and therefore, it improves the circuit bandwidth [82]. From a qualitative point of view, the inductor blocks the high-frequency current flowing through the resistor so that most of the bias current is solely used for charging and discharging the load capacitor, whereas the bias current is divided between the resistor and the load capacitor in the normal CS stage. Therefore, the ris-

ing and the falling times are reduced, indicating the improvement of the bandwidth. The quantitative details on the bandwidth enhancement of the shunt peaking is presented in [82], and it is proved that the maximum bandwidth enhancement ratio (BWER) of the shunt peaking is 1.84 with 1.5 dB peaking in magnitude response. The BWER is defined as the ratio of the 3-dB bandwidth with an inductive peaking over that without a peaking. On the other hand, it is slightly more difficult to understand the principle of the series-inductive peaking in the s-domain model as compared to the shunt peaking, because the inductor and the separation of the load capacitor introduce two additional poles but do not introduce any zero. Since there are three poles without zero, the bandwidth extension is dominated by the damping factor of the transfer function, which is not intuitively understood. Rather, a qualitative approach gives a much simple explanation. As shown in Figure 3.12(c), because the high-frequency current is blocked by the series inductor, *C1* and *C2* are charged sequentially. Practically, *C2* is charged after *C1* is fully charged. That is, the rise and fall times of the signal are reduced as a function of the ratio of *C1* to *C2*, at the cost of the increased delay [83]. The BWER of the series peaking can be higher than 2.5 with an optimum ratio of *C1* to *C2* [82].

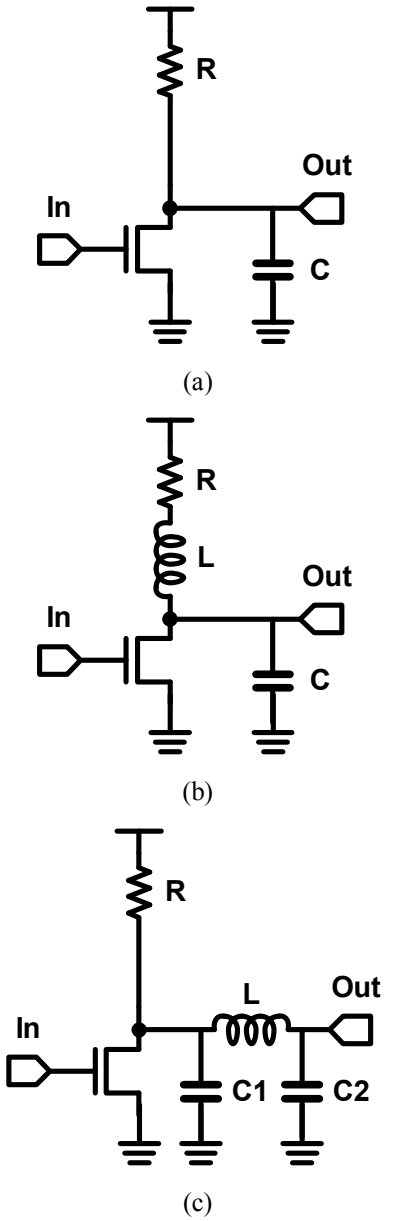


Fig. 3.12 Circuit diagrams of (a) basic CS stage, (b) CS stage with shunt peaking, and (c) CS stage with series peaking.

In Figure 3.13, some advanced inductive peaking techniques are illustrated. Figure 3.13(a) shows a shunt-and-series peaking which combines the shunt peaking and the series peaking. With a proper optimization, the shunt-series peaking can provide the BWER of 3.5 [81]. The inductive peaking technique shown in Figure 3.13(b) is widely referred to as a shunt-and-double-series peaking. By placing an additional inductor (*L3*) to the shunt and series peaking, it provides a better isolation of the capacitors, hence the circuit bandwidth can be further enhanced [83]. However, this technique requires three inductors, occupying extensive silicon area. A T-coil network shown in Figure 3.13(c) resolves this challenge [81], [83]. The negative coupling between two inductors (*L1* and *L2*) accounts for an initial boost in the current flow to *C2*, which means *C2* is effectively connected in series with the negative mutual inductance element of the T-coil [82]. That is, the T-coil network is equivalent to the shunt and double series peaking in Figure 3.13(b), where three inductors form a 'T' shape. That is why this technique is called a 'T-coil' network. With the T-coil network, the number of inductor is reduced to two. Moreover, the two inductors or the transformer can be implemented in silicon IC while occupying smaller area than that of two inductors without magnetic coupling, by using some novel winding techniques for T-coils [82], [83]. The achievable BWER of T-coil network is around 4 with a reasonable *C1* over *C2* ratio and gain peaking, and detailed quantitative values of BWER of the T-coil network are provided in [82].

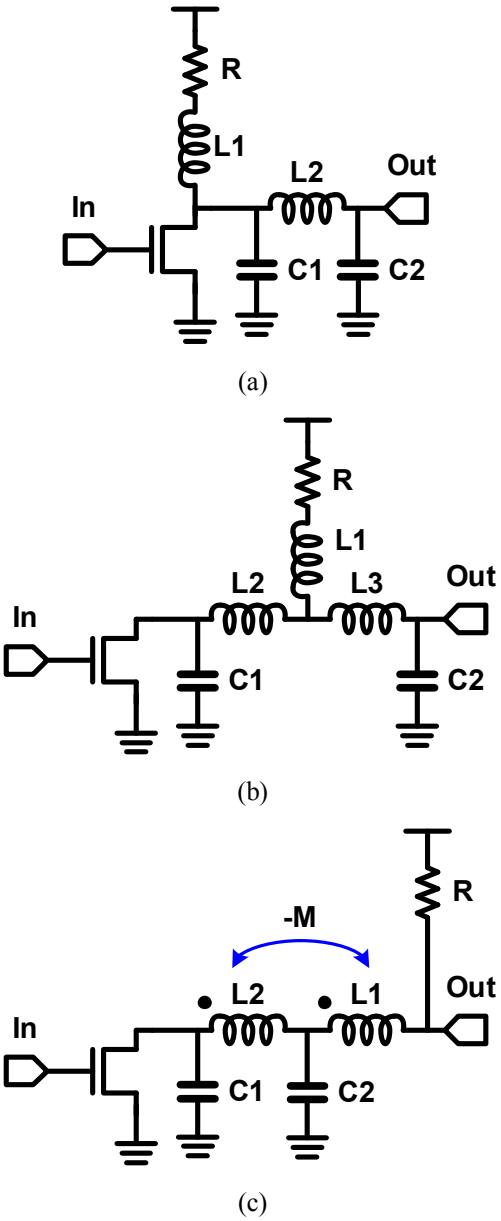


Fig. 3.13 Circuit diagrams of (a) shunt and series peaking, (b) shunt and double-series peaking, and (c) T-coil peaking.

From now on, the inductive peaking techniques employed for the TIA are introduced. In 2000, a Stanford group introduced a shunt inductive peaking to their TIA based on a resistive feedback CS stage as shown in Figure 3.14 [84]. A CG stage is placed between the shunt-peaked CS stage and the PD for decoupling the PD capacitance and the main stage and for introducing an additional flexibility to optimize the shunt inductor. With the proposed technique, they achieved BWER of 1.4 using a 20-nH on-chip inductor. Moreover, it is verified that their TIA achieves a 1.2-GHz bandwidth with 0.5-μm CMOS technology.

The first introduction of the series inductive peaking to the high-speed optical receiver was done by a Caltech research group in 2004 [66]. Their TIA, which was fabricated in 0.18-μm CMOS, consists of three gain stages in order to achieve a sufficient transimpedance gain and includes four series inductors. As shown in Figure 3.15(a), each series inductor separates adjacent capacitors as follows; 1) *L1*, the PD capacitance and the TIA input capacitance, 2) *L2* and *L3*, junction capacitances of

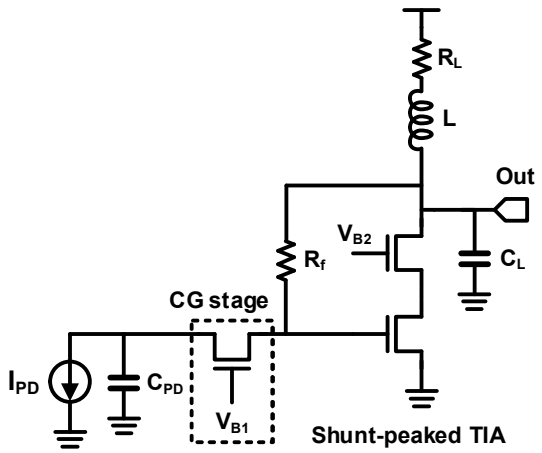
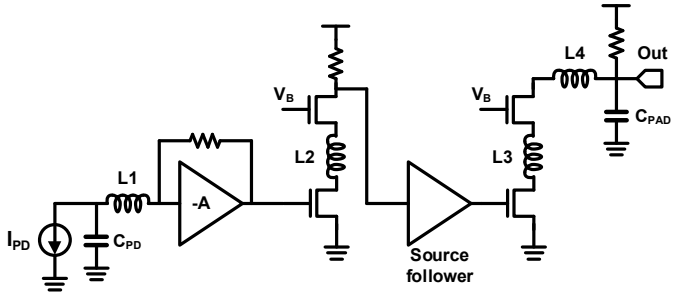


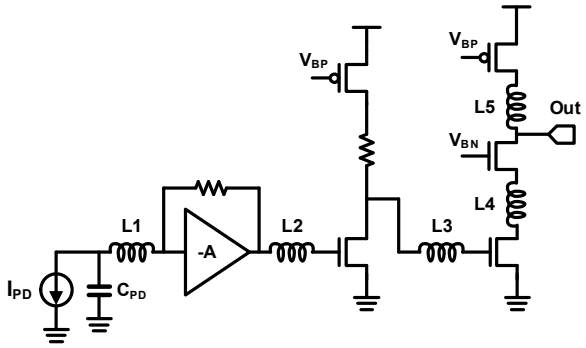
Fig. 3.14 TIA with shunt peaking presented in [84].

cascode transistors, and 3) *L4*, output parasitic capacitance and bonding pad capacitance, respectively. The four series inductors achieve the overall BWER of 2.4, and each contribution of the inductors is summarized in Figure 3.15(a). The 3-dB bandwidth of the TIA is 9.2 GHz and the eye diagram at 10 Gb/s is verified. In addition to using series inductors between cascade transistors, a UCSD group also employs series inductors between cascaded gain stages [85]. Two series inductors, *L2* and *L3*, are placed between cascaded amplifiers for further bandwidth extension while adopting a similar structure with [66] as shown in Figure 3.15(b). Moreover, it is notable that the authors of [85] did not overlook the degradation of phase linearity due to the increased BWER. They evaluated the group delay response of the TIA as well as the BWER, so the group delay variation of 16 ps is achieved for their TIA while that of the TIA in [66] exceeds 50 ps. As a result, although fabricated in 0.13-μm CMOS technology, the TIA opens the data eye at 40 Gb/s and achieves better eye opening compared to [66]. In 2012, the same group proposed to combine the series peaking technique with the inverter-based TIA as shown in Figure 3.15(c) [54]. From 1-V supply voltage, the TIA achieves a 30-GHz 3-dB bandwidth while dissipating 9 mW. They, again, emphasized the importance of optimizing the phase response, and they tried to suppress the group delay variation for better phase response. Therefore, the group delay variation is suppressed to less than 8 ps, thereby achieving a better eye opening compared to their previous work in [85].

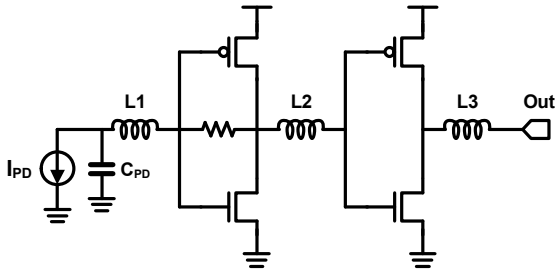


L used	BWER
L1	1.48
L2	1.42
L3	1.62
L4	1.17
L1, L2	2.0
L1, L2, L3	2.3
All	2.4

(a)



(b)



(c)

Fig. 3.15 (a) 3-stage TIA with series peaking [66], (b) examples of series peaking between cascaded gain stages presented in (a) [85], and in (c) [54].

T-coil peaking technique was introduced in a TIA design by Ehwa Womans University, in 2010 [86]. As shown in Figure 3.16, CG topology is chosen and two T-coil networks are employed to extend the bandwidth. The first T-coil network (T-coil1) separates C_{PD} from the junction capacitances of the transistors, and the second T-coil network (T-coil2) provides a shunt-peaking with R_L as well as separating the junction capacitance and the output load capacitance. The BWER of 2.3 and 3-dB bandwidth of 12.6 GHz are achieved, whereas the phase response is not fully optimized.

In addition to the inductive peaking techniques presented so far, an inductive feedback technique where an inductor is placed in series with a feedback resistor of the TIA can also be used. A straightforward intuition for the bandwidth extension using the inductive feedback is summarized in Figure 3.17, where R_F and L denote the feedback resistance and the inductance, respectively. Assuming a simple CS amplifier whose gain and 3-dB bandwidth are $-g_m R_{out}$ and f_p, respectively, a resistive feedback extends the bandwidth by a factor of R_{out}/R_L, where R_L is the resistance of the parallel combination of R_{out} and R_F, while the gain is sacrificed by the same fac-

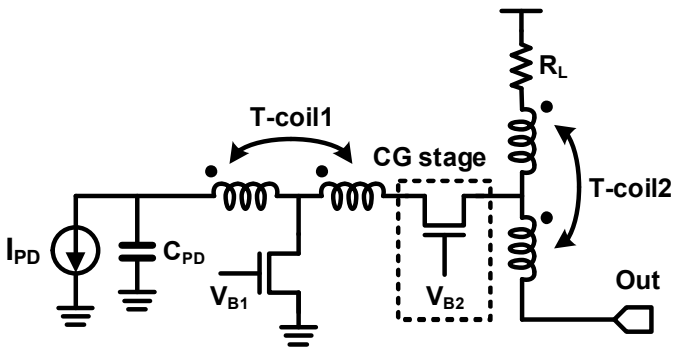
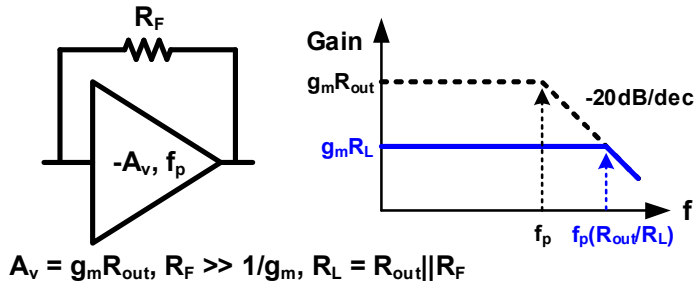
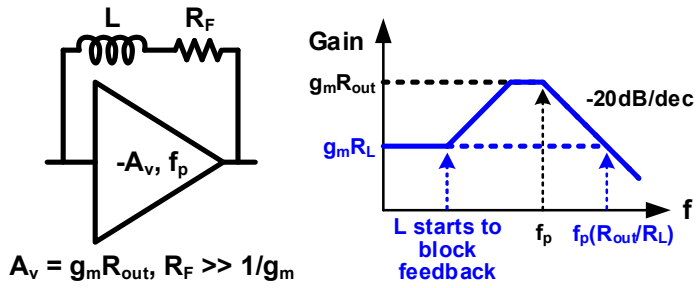


Fig. 3.16 CG TIA with T-coil peaking [86].

tor, as shown in Figure 3.17(a). On the other hand, insulting a series inductor to R_F, as shown in Figure 3.17(b), introduces a peaking as well as the bandwidth extension. At low frequencies, because the impedance of the inductor is almost zero so that it can be neglected, the transfer function with the inductor becomes the same as that without the inductor. However, the impedance of the inductor increases as the frequency increases hence becomes larger than that of the R_F at a certain frequency, which can be considered as a zero frequency. It means that the feedback strength



(a)



(b)

Fig. 3.17 Effect of feedback on transfer function of amplifier (a) with only resistive feedback and (b) with combination of resistive and inductive feedback.

decreases, and therefore the transfer function begins to follow that without the resistive feedback. To be simple, when the inductance dominates the overall feedback impedance, the feedback network becomes negligible. That is, the transfer function exhibits a peaking as shown in Figure 3.17(b), which follows the low-gain curve at low frequencies and the high-gain curve at high frequencies. The inductive feedback technique will be revisited with a detailed analysis in the later chapter. Luxtera adopted the inductive feedback technique into the CS-stage-based TIA in 2006 [29]. They also reported that the inductor reduces the input-referred noise by the feedback resistor. In [87] and [88], an inductive feedback was employed in the inverter-based TIA, in order to compensate for the frequency-dependent loss due to the PD bonding parasitic elements. With the inductive peaking, the 3-dB bandwidth of the TIA is extended from 11.4 GHz to 25.2 GHz, which corresponds to a BWER of 2.2 dB.

There in another inductive peaking technique proposed by TSMC in 2014, which is called a shared inductor [71]. By sharing an inductor by adjacent amplifier stages, the number of inductors used for the entire chip halves, and moreover, the size of each inductor can be further reduced. It was reported that the overall chip area is reduced by 56 % and the power consumption is also saved by 27 % compared to the case with the conventional inductive peaking. Since the on-chip inductor occupies a huge area in ICs, the shared inductor is a good approach because it can dramatically reduce the chip area.

3.3.2 Equalization

For the receivers used in the electrical link, various types of equalization techniques have been utilized to overcome the frequency-dependent loss of the electrical transmission line. On the other hand, there has been little interest in using such equalization techniques for optical links because optical fibers provide an almost infinite transmission bandwidth. However, recent state-of-the-art optical receivers began to include equalization techniques in order to 1) overcome the limit of the O/E conversion bandwidth, 2) eliminate the tradeoff between the gain and the bandwidth, 3) enhance the TIA bandwidth with minimal added noise. This section summarizes the equalization techniques for the optical receiver implemented so far.

In 2010, a group in University of Toronto proposed a TIA with negative Miller capacitances and a continuous-time linear equalizer (CTLE) [89], in order to overcome the limited speed of a spatially modulated PD. In [89], a pair of capacitors are placed between the input and the output of a differential CS-stage TIA, which have the same polarity, as shown Figure 3.18. Due to the Miller effect, the input capaci-

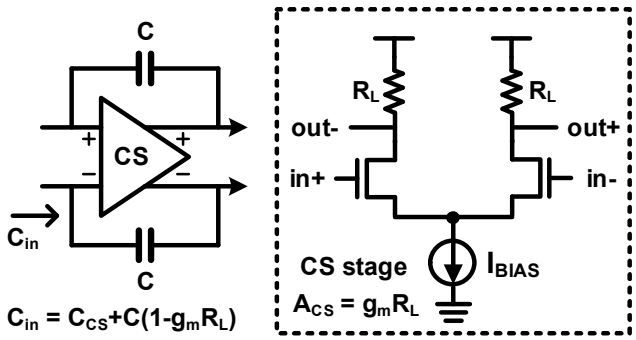


Fig. 3.18 Miller capacitance applied to differential CS stage [89].

tance of the TIA becomes $C_{cs}+C(1-A_{cs})$, where C_{cs}, C, and A_{cs} are the input capacitance of the CS stage, the Miller capacitance, and the gain of the CS stage, respectively. Note that the second capacitance term from the Miller capacitance has a negative value when the gain of the CS stage is larger than unity. That is why this technique is referred as the negative Miller capacitance. That is, the input capacitance becomes lower than C_{CS}, therefore extending the bandwidth of the preceding amplifier. A CTLE introduces a similar effect with the inductive feedback technique discussed in the previous section. A CTLE is generally implemented with the CS stage with a source degeneration using the combination of a resistor and a capacitor. As shown in figure 3.19, the source degeneration with only a resistor decreases the gain of the CS stage by a factor of $1+g_mR_s$ while the circuit bandwidth is extended by the same factor. By placing a degeneration capacitor (C_s) in parallel with R_s, the degeneration strength becomes weakened at high frequencies so that the transfer function follows that of the normal CS stage, which is similar to the effect of the inductive feedback. As a result, a CTLE achieves a high-frequency peaking in its transfer function as shown in Figure 3.19. The overall architecture presented in [89] is described in figure 3.20. Two identical CS stages are cascaded for a high gain, and the negative Miller capacitance is used for the second stage to reduce the load capacitance driven by the first CS stage. After the cascaded CS stage, a CTLE is used to compensate for the high-frequency loss due to the limited bandwidth of the spatially modulated PD, which is monolithically integrated on the same chip. Therefore, [89] achieves a 5-Gb/s operation while the 3-dB bandwidth of the spatially modulated PD is only 700 MHz.

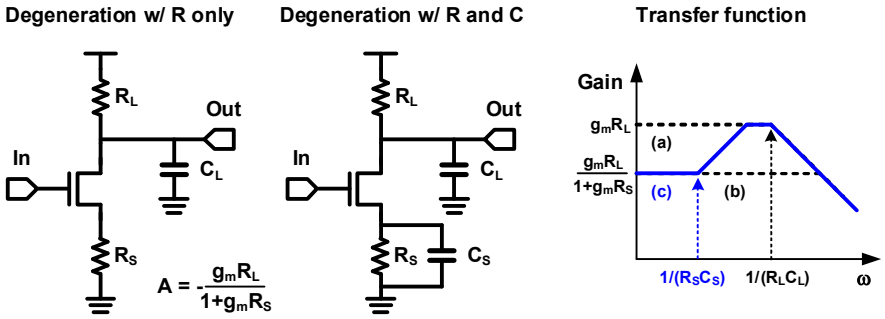


Fig. 3.19 Basic principle of CTLE.

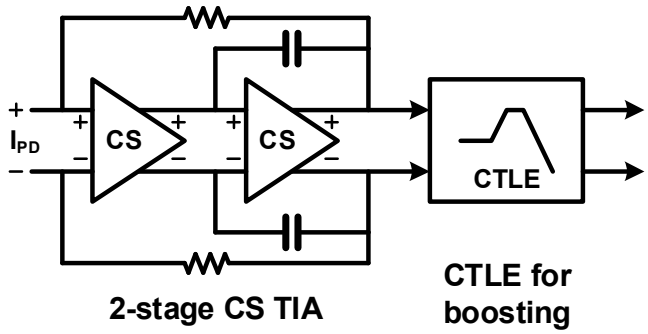


Fig. 3.20 Overall architecture of optical receiver [89].

In order to achieve a sufficient high-frequency boosting and to avoid noise amplification of the CTLE, a nonlinear equalization technique such as a decision feedback equalizer (DFE) can be employed. The DFE technique has been widely used in electrical links for compensating for an inter-symbol interference (ISI) that comes from the frequency-dependent loss and reflections of the electrical channel. Because of the loss and the reflections, a single transmitted bit cannot complete its transition within a bit period or a unit interval (UI) so that it influences the following bit

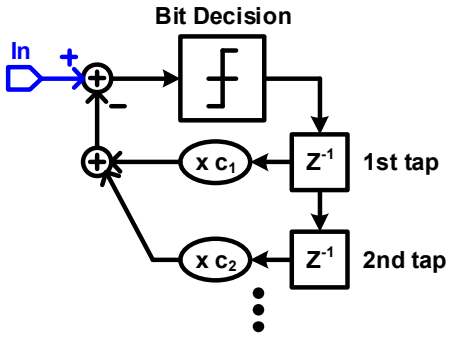


Fig. 3.21 Simplified block diagram of DFE.

sequence. The basic concept of the DFE is that the previously received bit sequence is used when a receiver make a decision whether the currently received bit is '0' or '1', by effectively cancelling the influence of the previous bit as described in Figure 3.21. That is, the DFE uses consecutive multiple bits to make a decision for a single bit because the bits interact with one another due to the ISI. The degree of the compensation is closely related to the number of the DFE taps, which determines the length of the previous bit sequence to be used for the decision of the current bit. However, the power consumption of the DFE circuitry is proportional to the number of the taps. It means that DFE needs a huge power to achieve a sufficient compensation, which makes it less attractive to adopt the DFE for optical links.

In 2013, however, IBM proposed an optical receiver employing a DFE with infinite impulse response (IIR) feedback [55]. An interesting feature is that the receiver uses a resistor-based TIA as shown in Figure 3.22. As mentioned earlier, the resistor-based TIA should sacrifice the bandwidth for the gain and the SNR. In this approach, a large resistor is used for the large gain and SNR at the cost of the reduced

bandwidth, and the resulting ISI is compensated by the DFE. Moreover, only one DFE tap is used to compensate the ISI which comes from nearly infinite number of preceding bits by using an IIR feedback, and therefore the overhead of a large number of the DFE taps is dramatically reduced. In contrast to the electrical link where the ISI is mainly caused by the external components so that they are not easy to predict, the RC time constant, mainly arising from the PD capacitance and the TIA resistance, is truly predictable. Thus, by making the feedback path experiences the same RC time constant as the input signal, it can compensate for the ISI as shown in Figure 3.22. With the IIR-DFE technique, the resistor-based TIA achieves a 9-Gb/s operation. Recently, [90] extends the adoption of the IIR-DFE to the RGC TIA structure where the dominant pole is located at the output of the TIA. Compared to [55], the dominant pole is moved from the TIA input to the output, therefore the feedback summer is placed at the output of the TIA. The data rate of 20-Gb/s is achieved without using any inductive peaking techniques with a relatively small area of 0.027 mm^2.

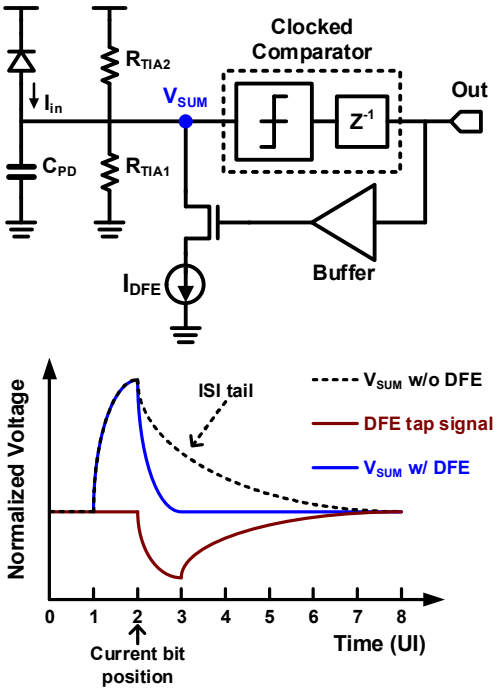


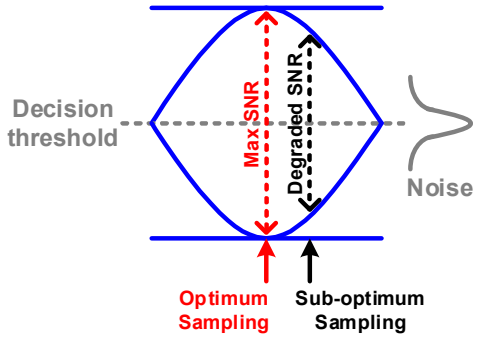
Fig. 3.22 Resistor-based TIA with IIR-DFE [55].

3.4 Clock and Data Recovery Circuits

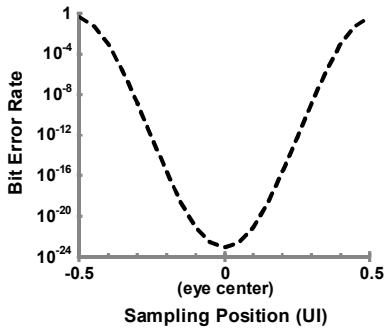
3.4.1 CDR Basic

Because the transmission of binary data in optical communications is based on a time-division multiplexing, a simple amplification of the incoming signal is not sufficient to recover the data. The other requirement for the complete recovery of the

data is to retime the incoming data with a clock whose frequency is well matched to the data rate. In general, the retiming operation is based on a sample-and-regeneration using a clocked sense amplifier. That is, the timing information of the data is recovered by sampling, while the value of the binary data is gathered by sampling and is recovered to a digital rail-to-rail swing by regeneration. Note that the SNR at the sense amplifier input is restored after being regenerated at the sense amplifier output, unlike linear amplifiers where the input SNR propagates to the



(a)



(b)

Fig. 3.23 Dependency of BER on sampling time. (a) Qualitative explanation of BER degradation and (b) quantitative relation of BER on sampling time [91].

output, because the regeneration is based on the positive feedback. Rather, the input SNR leads to a probability of a wrong decision while the output has little noise component. Therefore, during the retiming operation, it is important to sample the data at a proper time where the SNR is maximized. Unlike the retiming of the digital signal where only the setup and the hold times are critical, the error rate in case of the retiming of the analog signal varies with the timing, even though the setup and the hold times are sufficient. Figure 3.23(a) illustrates this aspect. In [91], the theoretical dependency of the bit-error-rate (BER) over the sampling time is analyzed. The calculated BER as a function of the sampling time using the formula presented in [91] is shown in Figure 3.23(b), where the maximum SNR of 10 is used for the calculation. For simplicity, it is assumed that there is no random timing error. As observed in Figure 3.23(b), the BER is exponentially degraded as the sampling time deviates from the ideal position. A clock and data recovery (CDR) circuit generates the sampling clock with a precise frequency and drives the sampling time to the ideal position. In conventional optical receivers, the importance of the CDR was underestimated or even ignored for a long time. However, recently, there have been some engineering examples which include the CDR circuits for better performances and robustness. This section provides examples of the CDR circuit suitable for a high-speed optical receiver.

3.4.2 CDR Examples

The first CDR example to be introduced is presented in [29]. As shown in Figure 3.24, it consists of two negative feedback loops; one is for the precise phase align-

ment and the other is for the initial frequency acquisition. To avoid the interference between the loops, a locking procedure that locks the loops sequentially is used in the CDR. In the initial state, the frequency acquisition loop is enabled until the frequency lock detector decides that the frequency of the CDR clock is sufficiently close to the data rate using the external reference frequency. After that, the CDR enters the normal CDR operation, where the lock detector activates the phase alignment loop and simultaneously, disables the frequency acquisition loop, in order to track the optimum sampling phase. Although the frequency acquisition loop is disabled in the normal operation, the residual frequency error can be tracked only with the phase alignment loop, as long as the frequency error is within the lock-in range. If a large frequency error outside the lock-in range is introduced, the lock detector enables the frequency acquisition loop again. Note that a single voltage-controlled oscillator (VCO) generating the CDR clock is shared between the two loops. Because the clock frequency generated from the VCO is proportional to the control

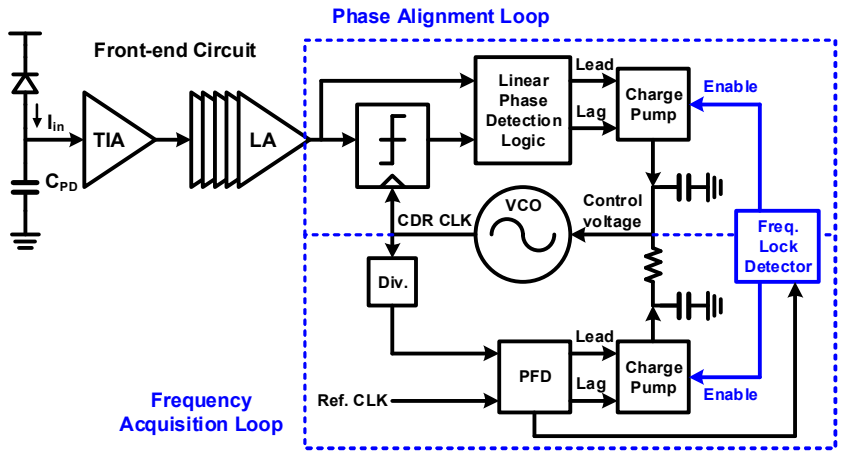


Fig. 3.24 Sequential locking CDR presented in [29].

voltage of the VCO, the phase of the clock is adjusted by changing the control voltage of the VCO for a short period of time, while the control voltage is changed permanently for the frequency acquisition. In the front-end, the limiting amplifier restores the low-swing signal from the TIA to a logic level. It is because this CDR uses the linear phase detection scheme which requires a logic processing of the data before entering into the recovery. The linear phase detection scheme provides a predictable loop dynamics of the CDR compared to the nonlinear phase detection scheme which is also known as a bang-bang phase detection or an Alexander phase detection. However, a huge power is dissipated in this CDR structure, due to the limiting amplifier stage and the logic gate operating at a high data rate. With the implementation of the CDR, the receiver achieves a remarkable sensitivity of -19.5 dBm at 20 Gb/s. Similarly, [63] achieves a 25-Gb/s operation except that it uses a mixer-based phase detector instead of the conventional linear phase detector to improve the speed.

The next CDR example is described in Figure 3.25 [79]. Unlike the sequential locking CDR, it adopts a dual-loop CDR architecture where the phase alignment and frequency acquisition loops operate simultaneously. In this type of CDR, the loop bandwidth of one loop should be set an order of magnitude lower than that of the other loop, in order to prevent the interference between the loops. In this work, the loop bandwidth the phase alignment loop is designed to be lower than that of the frequency acquisition loop to suppress the jitter transfer caused by the input data. Contrary to the sequential locking CDR, the VCO is not shared between the loops. An additional phase adjusting circuit based on a phase interpolator (PI) is used for the phase alignment loop. The PI cannot generate a clock but can synthesize a phase

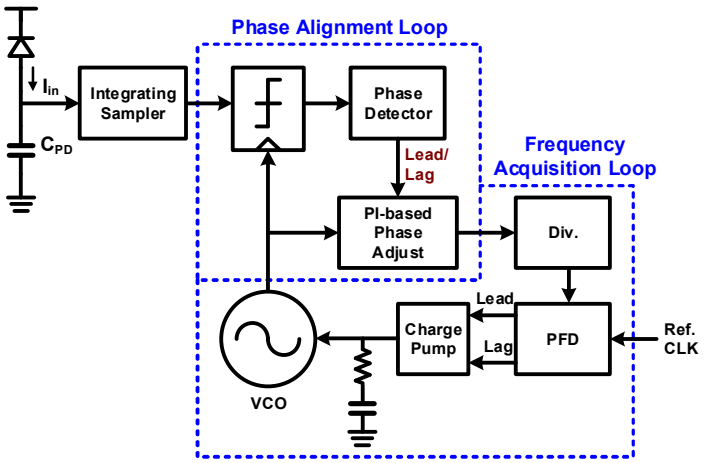


Fig. 3.25 Dual-loop CDR with integrating front-end [79].

using the clock generated by the VCO. The phase-adjusted clock is fed back to the VCO through a phase-frequency detector (PFD) so that the output phase of the VCO is eventually aligned with the data. It achieves a sensitivity of -5.4 dBm at 16 Gb/s while consuming 58 mW in total.

As CMOS technology scales down, a lot of non-idealities appear in analog circuits, degrading the performance. In order to overcome the performance degradation introduced by the analog circuits such as a loop filter and a charge pump, a recently developed all-digital CDR substitutes a digital loop filter for the analog loop filter and the charge pump [75]. For the all-digital implementation, the analog VCO also should be replaced by a digitally-controlled oscillator (DCO) to interface with the digital loop filter. On the other hand, the main drawback of this all-digital approach is the long loop latency, which introduces a phase dithering and a loop stability issue. To alleviate this problem, a direct phase-control path is enabled in [75], achieving a

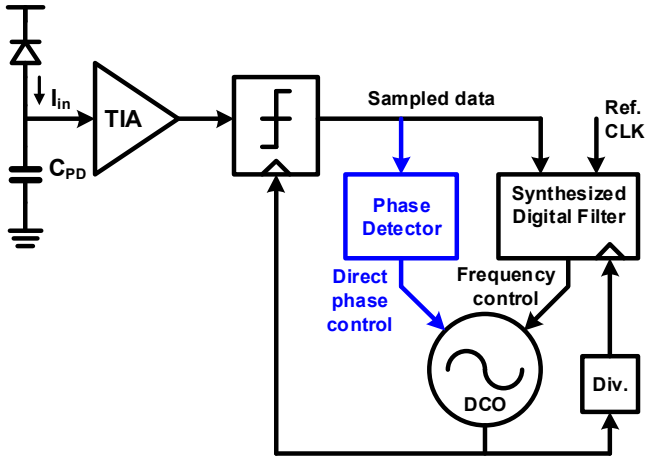


Fig. 3.26 All-digital CDR presented in [75].

quick phase response as shown in Figure 3.26. The implemented chip operates at the data rate of 26.5 Gb/s, consuming 254 mW. It is noteworthy that the all-digital approach is surely promising as the CMOS technology continues to scale down, making the digital performance better and better.

Chapter 4

Low-Power Optical Receiver Front-End

4.1 Overview

As previously discussed, a CMOS implementation of the optical receiver at high speed is essential to fully take advantage of silicon photonics. In this regard, we propose a high-speed and low-power optical receiver front-end with a target data rate of 20 Gb/s. We provide a full description on circuit techniques and design procedures, particularly focusing on the techniques for a bandwidth extension.

In this chapter, we present some theories of the TIA topology employed for this receiver first. Then, detailed circuit implementations will be given, followed by measurement results of the prototype chip.

4.2 Inverter-Based TIA with Resistive Feedback

In this section, we describe the previously discussed inverter-based TIA in more detail. A circuit diagram and its equivalent small signal including input and output load conditions are drawn in Figure 4.1. The advantage of this topology manifests itself in this model. First, it features a very simple architecture which consists of only two transistors and does not need any bias circuits, thanks to which it best suits to modern CMOS technologies. Secondly, in contrast to conventional TIA topologies, it uses both the PMOS and the NMOS, thereby providing higher g_m with low-power consumption. These distinct merits have enabled a widespread use of this topology. Here, we provide some useful results from numerical analysis of this topology.

In general, the P/N ratio of the inverter is set to 2–3 for the compensation of unequal electron and hole mobility values, which is generally applied to logic inverters to ensure symmetric rising and falling transitions. However, when it comes to an

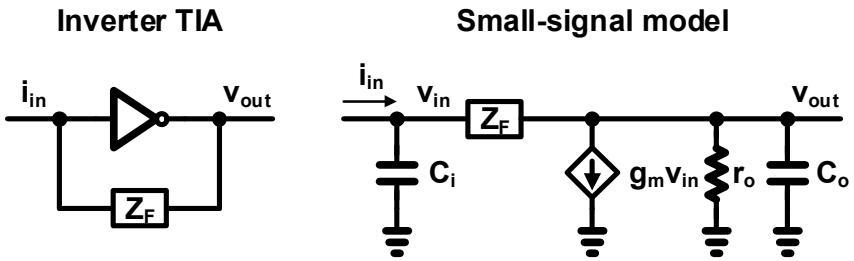


Fig. 4.1 Inverter-based TIA.

amplifier, such a criterion is not valid anymore, and an alternative criteria should be established to optimize the amplifier performance. In fact, similar approaches have been presented with respective optimization goals. In [92] and [93], the relation between the P/N ratio and the metastability resolution time constant was revealed. And [94] provides a similar result, but with a purpose of optimizing the inverter-based driver. Likewise, an intrinsic bandwidth, ω_i, of the inverter amplifier is defined here, as

$$\omega_i = \frac{g_m}{C_{in}} = \frac{g_{mp} + g_{mn}}{C_{ox}\left(W_p + W_n\right)L}, \tag{4.1}$$

where g_{mp} and g_{mn} are the transconductances of the PMOS and the NMOS, respectively, and C_{in} is the total gate capacitance of the transistors. If the P/N ratio is chosen in a way that maximizes ω_i, we can achieve targeted g_m with the minimum transistor sizes.

An optimum P/N ratio to maximize ω_i can be found as follows. With the supply voltage of V_{DD} given and alpha-power law applied to the current equations, the drain currents of the PMOS and the NMOS operating in the saturation region are given by,

$$\begin{aligned} I_p &= \frac{1}{2}k_p C_{ox}\frac{W_p}{L}\left(V_{DD} - V_G - V_{THP}\right)^{\alpha} \\ &= \frac{1}{2}k_p C_{ox}\frac{rW_n}{L}\left(V_{DD} - V_G - V_{THP}\right)^{\alpha} \end{aligned} \tag{4.2}$$

and

$$I_n = \frac{1}{2}k_n C_{ox}\frac{W_n}{L}\left(V_G - V_{THN}\right)^{\alpha}, \tag{4.3}$$

where k_p and k_n are the fitted constants for the alpha-power, r is the P/N ratio, and V_G is the gate voltage. With no bias current applied to the input, the currents in (4.2)

and (4.3) are the same, from which V_G can be expressed as a function of r as follows.

$$V_G = \frac{k_p^{1/\alpha} r^{1/\alpha} \left(V_{DD} - V_{THP}\right) + k_n^{1/\alpha} V_{THN}}{k_p^{1/\alpha} r^{1/\alpha} + k_n^{1/\alpha}}. \tag{4.4}$$

Rearranging (4.2) and (4.3) with (4.4) results in

$$V_{DD} - V_G - V_{THP} = \frac{k_n^{1/\alpha} \left(V_{DD} - V_{THP} - V_{THN}\right)}{k_p^{1/\alpha} r^{1/\alpha} + k_n^{1/\alpha}} \tag{4.5}$$

and

$$V_G - V_{THN} = \frac{k_p^{1/\alpha} r^{1/\alpha} \left(V_{DD} - V_{THP} - V_{THN}\right)}{k_p^{1/\alpha} r^{1/\alpha} + k_n^{1/\alpha}}. \tag{4.6}$$

Meanwhile, the transconductances of the PMOS and the NMOS can be written as

$$g_{mp} = \frac{\alpha}{2} k_p C_{ox} \frac{r W_n}{L} \left(V_{DD} - V_G - V_{THP}\right)^{\alpha - 1} \tag{4.7}$$

and

$$g_{mn} = \frac{\alpha}{2} k_n C_{ox} \frac{W_n}{L} \left(V_G - V_{THN}\right)^{\alpha - 1}. \tag{4.8}$$

After substituting (4.5) and (4.6) into (4.7) and (4.8), ω_i (defined above) can be obtained by definition and written by

$$\begin{aligned} \omega_i = \frac{g_m}{C_{in}} &= \frac{g_{mp} + g_{mn}}{C_{ox}\left(1 + r\right) W_n L} \\ &= K_{cont} \frac{r^{1 - 1/\alpha}}{1 + r} \left(K_p^{1/\alpha} r^{1/\alpha} + K_n^{1/\alpha}\right)^{2 - \alpha} \end{aligned}, \tag{4.9}$$

where K_{cont} is the product of the remaining constants. On the other hand, the solution of the following equation,

$$\frac{\partial \omega_i}{\partial r} = 0, \tag{4.10}$$

gives an optimum P/N ratio. Interestingly, if α is given by 2 for conventional long-

channel devices, (4.9) is reduced to

$$\omega_i = K_{cont} \frac{\sqrt{r}}{1+r}. \tag{4.11}$$

It indicates that ω_i has the maximum value when r is equal to 1, meaning when the widths of the PMOS and the NMOS are exactly the same. However, for short-channel devices where α is lower than 2, the optimum P/N ratio slightly deviates from 1. Due to the complexity of solving the equation, an empirical approach would be more straightforward. Based on the simulations, it is found that α is about 1.3 and k_p is about two times smaller than k_n. The optimum P/N ratio empirically obtained is 0.7 with only 1 % improvement of ω_i, as compared to the case when r is 1. This means that it is still reasonable to choose an optimum ratio as 1, especially when considering that the common level, V_G could decrease too much with small r. This result is well supported also by the simulation results indicating that ω_i has a peak value when the P/N ratio is around 1 regardless of the supply voltages as shown in

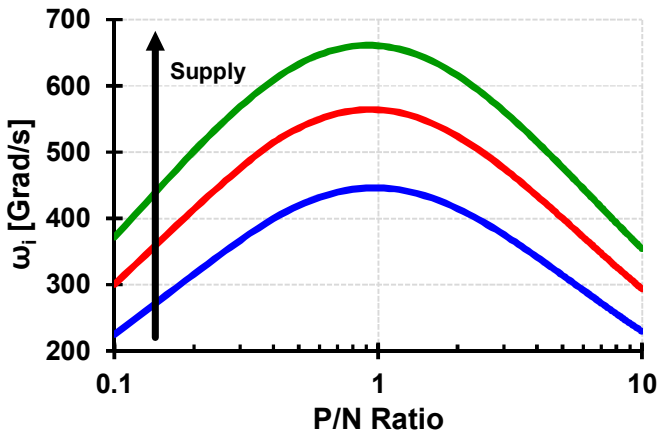


Fig. 4.2 Simulated ω_i with respect to P/N ratio for various supply voltages.

Figure 4.2 [94].

On the other hand, the transfer function of the inverter-based TIA can be obtained from the small-signal model in Figure 4.1 and is given by

$$R_T = \frac{v_{out}}{i_{in}} = \frac{R_o\left(1 - g_m Z_F\right)}{1 + sR_o\left[\left(1 + Z_F / r_o\right)C_i + C_o\right] + s^2 Z_F R_o C_i C_o}, \tag{4.12}$$

where R_o is the output resistance of the TIA and can be written as

$$R_o = \frac{r_o}{1 + g_m r_o} \approx \frac{1}{g_m}. \tag{4.13}$$

If the feedback element is purely resistive with R_F, (4.12) becomes

$$R_T = \frac{R_o\left(1 - g_m R_F\right)}{1 + sR_o\left[\left(1 + R_F / r_o\right)C_i + C_o\right] + s^2 R_F R_o C_i C_o}. \tag{4.14}$$

With the second-order term in the denominator neglected, the bandwidth of the TIA can be approximated as

$$f_{-3dB} \approx \frac{1}{2\pi R_o\left[\left(1 + R_F / r_o\right)C_i + C_o\right]}. \tag{4.15}$$

Here, we assume that C_i is dominantly determined by a large PD capacitance. Note that the bandwidth is highly dependent on R_o, which means that increasing g_m can enhance the bandwidth while the transimpedance gain gets closer to R_F. The overall characteristic is quite similar to that of the CG-based TIA (in Chapter 3) except that g_m of the inverter TIA can be much higher with the same power budget, which is a distinguished advantage of this topology. The only way to improve g_m is increasing the inverter size with the fixed supply voltage. Thus, we can expect that the bandwidth would increase as the size of the inverter is increased until the input and the output capacitances of the inverter become comparable to the PD capacitance as shown in Figure 4.3. If the capacitances of the inverter begin to dominate C_i, in-

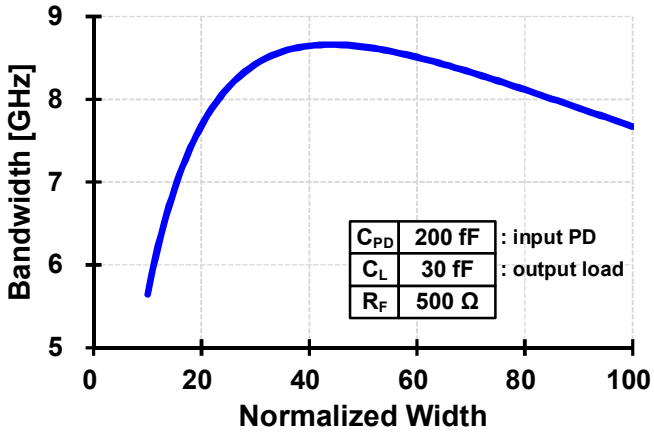


Fig. 4.3 Simulated bandwidth with increasing size of inverter.

creasing the size has little effect on the bandwidth and the bandwidth will saturate. And then, the bandwidth will start to decrease as R_F/r_o in (4.14) becomes larger, which is clearly observed in Figure 4.3.

To make matters worse, though not included in the initial model, the gate-drain capacitance of the transistor also degrades the bandwidth. With this effect taken into consideration, Z_F in (4.12) should be replaced by the parallel combination of R_F and $1/sC_F$, resulting in

$$R_T = \frac{v_{out}}{i_{in}} = \frac{R_o\left(1-g_m R_F\right)\left[1-sR_F C_F/\left(g_m R_F-1\right)\right]}{1+s\left[R_F C_F + R_o\left[\left(1+R_F/r_o\right)C_i + C_o\right]\right]+\cdots}. \tag{4.16}$$

Now, we have an additional first-order term in the denominator, and a zero on the left-half plane (LHP) is introduced. Again, higher-order terms are not taken into consideration. As the effect of sR_FC_F directly shows up in the transfer function, the bandwidth saturation observed in Figure 4.3 would shift forward, severely restricting

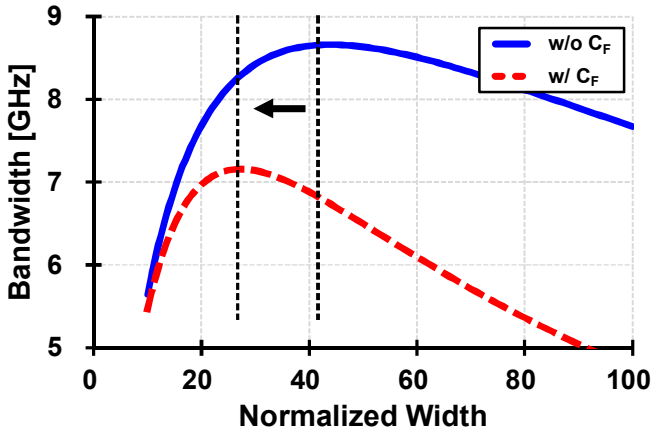


Fig. 4.4 Comparison between the cases without C_F and with C_F.

the maximum available bandwidth as described in Figure 4.4. It is also important to assess the effect of the LPH zero. From (4.16), the zero frequency is given by

$$f_{z,LHP} = \frac{g_m R_F - 1}{R_F C_F} \approx \frac{g_m R_F}{R_F C_F} = \frac{g_m}{C_F}. \tag{4.17}$$

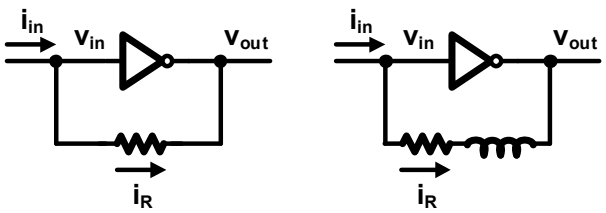
In general cases where g_m/C_F is much higher than $1/(R_F C_F)$, the effect of the zero is negligible. However, a poor design (i.e. the case when $g_m R_F \approx 1$) may make the zero frequency significantly close to the pole frequency, hence degrading the overall frequency response, especially the phase response.

From these observations, we can conclude that the bandwidth of the inverter-based TIA is severely limited in nature, which cannot be overcome by consuming more power. The only solution is to reduce R_F, hence the gain, resulting in the degradation of noise figure.

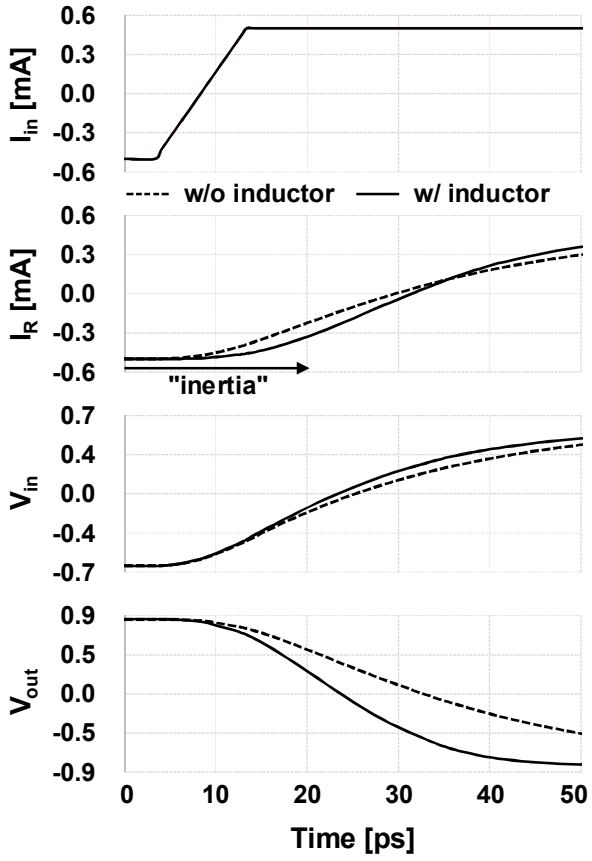
4.3 Inverter-Based TIA with Resistive and Inductive Feedback

As briefly discussed in Chapter 3, the inductor is most popularly used to improve the bandwidth of the amplifier, because it provides the highest bandwidth extension ratio (BWER) with no additional power penalty. However, the inductor-peaking techniques have been rarely employed for the inverter-based TIA except the case where the inductor-series peaking was used [54]. Although the inductor-series peaking improves the bandwidth, the BWER is strictly limited by the capacitive load conditions. Hence, using the series-peaking technique alone is not sufficient for the high BWER. Alternatively, the inductor-shunt peaking can also be an effective technique for the inverter-based TIA [87]. In this regard, we present a detailed analysis of the shunt-peaking technique that is specific to the inverter-based TIA.

To start with, the effect of the shunt-peaking technique can be predicted in a qualitative view. The inverter-based TIAs without and with an inductor are shown in Figure 4.5(a) and time-domain simulation results are presented in Figure 4.5(b). As shown in Figure 4.5(b), the event begins with a fast current ramp applied to the input. Here, we assume that the input rise time is considerably shorter than the time constant of the TIA without the inductor in order to clearly show the difference. At the beginning, both input nodes start to be charged by the same amount. The difference between the two cases becomes clear as the input voltages are increased slightly.



(a)

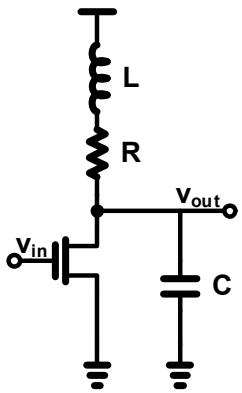


(b)

Fig. 4.5 (a) Inverter-based TIA without and with inductor. (b) SPICE verification of improved frequency response with inductor.

In the case without the inductor, the increased input voltage is instantaneously reflected to the feedback current, I_R, resulting in a decrease of I_R, or the current flowing into the output node is increased. Hence, charging the input node and discharging the output node are both retarded significantly, degrading the bandwidth. On the other hand, in the case of employing the inductor, the increased input voltage does not induce a change in I_R due to the property that an inductor is reluctant to change its current. As a result, both the input and the output nodes are charged and discharged more quickly, which is not hindered by the feedback current for the time being.

This phenomenon can also be understood in a quantitative manner. Before getting into the analysis, it can be instructive to find an analogy with the conventional shunt-peaking technique. The basic principle and some useful formula of the shunt-peaking technique are summarized in Figure 4.6 [84], which shows an example of the shunt-peaking technique applied to the common-source amplifier. In [84], the



$$\frac{v_{out}}{v_{in}} = \frac{g_m(R+sL)}{1+sRC+s^2LC}$$

$$m = \left(\frac{L}{R}\right)/(RC) = \frac{L}{R^2C}$$

m	ω_{-3dB}	Response
0	1.00	No shunt peaking
0.32	1.60	Optimum group delay
0.41	1.72	Maximally flat
0.71	1.85	Maximum bandwidth

Fig. 4.6 Example of inductor-shunt peaking [85].

ratio of the *L/R* and the *RC* time constants is defined as *m*, and *m* is used as a performance metric for inductor-shunt peaking, which is summarized in Figure 4.6. From this metric, one can carefully choose design parameters for the optimum performance. However, the usage of the metric presented in [84] is limited only to a certain amplifier type, necessitating another strategy for the inverter-based amplifier.

In general, an amplifier with no peaking technique can be approximated as a second-order system and its transfer function can be written as

$$H(s)=\frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}, \tag{4.18}$$

where ω_n is the natural frequency and ζ is the damping factor. It is well known that the shape of the magnitude response entirely relies on ζ and can be categorized into the following:

$$\begin{cases} \zeta > 1 & \text{(Overdamped)} \\ \zeta = 1 & \text{(Critically damped)} \\ \zeta < 1 & \text{(Underdamped)} \\ \zeta = 0 & \text{(Undamped).} \end{cases}$$

In Figure 4.7, the magnitude response of each case is drawn. The underdamped case with $\zeta = 1/\sqrt{2}$, represents the point where the response is maximally flat while, with $\zeta < 1/\sqrt{2}$, a significant peaking is observed.

On the other hand, an amplifier with an inductor introduced typically exhibits a transfer function in the form of (4.18) with additional zero. In this case, the transfer function can be written by

$$H(s)=\frac{1+s/\omega_z}{1+s\cdot 2\zeta/\omega_n+s^2\cdot(1/\omega_n)^2}, \tag{4.19}$$

where ω_z denotes the zero frequency. Decomposing the transfer function into two

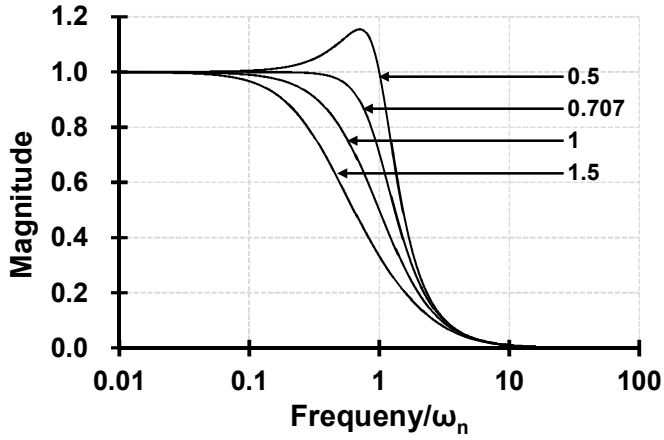


Fig. 4.7 Magnitude responses with various damping factors.

parts and assessing them separately can be helpful. It is notable that the newly introduced zero can never bring the effect of bandwidth extension for the case with $\zeta < 1/\sqrt{2}$. This observation can be a good criteria for the effectiveness of peaking techniques. The ratio, *m* defined in [84] may be still useful, because it approximately estimates the degree of peaking or can be directly translated into the BWER. From (4.19), *m* can be redefined as

$$m = \frac{1}{2\zeta} \cdot \frac{\omega_n}{\omega_z}. \tag{4.20}$$

It can be generalized by obtaining each BWER with respect to *m* for various damping factors as visualized in Figure 4.8, from which the BWER can be roughly estimated. In summary, the introduction of an inductor would be applicable only to highly-damped cases or should have little effect on the damping factor.

The example in Figure 4.6 can be illustrated using the redefined formula. The transfer function can be rewritten in the form of (4.19), where

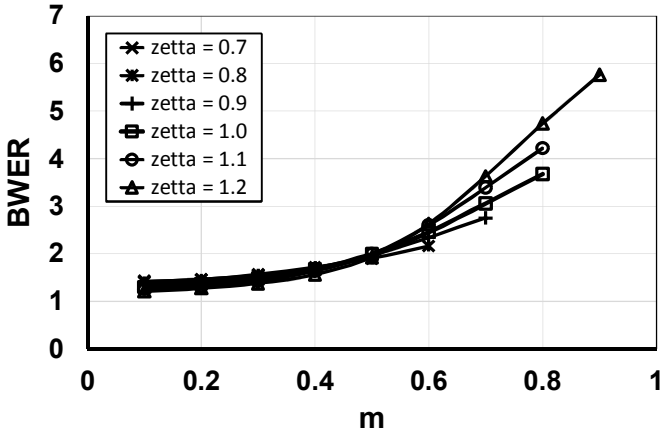


Fig. 4.8 BWER with respect to *m* for various values of ζ.

$$\omega_n = 1/\sqrt{LC}\,, \tag{4.21}$$

$$\omega_z = R/L\,, \tag{4.22}$$

and

$$\zeta = \frac{1}{2}R\sqrt{\frac{C}{L}}\,. \tag{4.23}$$

From (4.21), (4.22), and (4.23), the relation between ζ and *m* can be obtained and give by

$$\zeta = \frac{1}{2\sqrt{m}}\,. \tag{4.24}$$

Interestingly, ζ and *m* trade off with each other, that increasing *m* results in the decrease of ζ. Particularly, ζ reaches the critical value of 0.707 when *m* becomes 0.5, meaning the maximally available *m* is limited to 0.5, thereby also restricting the maximum BWER of the conventional shunt-peaking technique.

Similarly, the BWER of the inductive feedback applied to the inverter-based TIA

can be estimated. By substituting R_F+sL_F for Z_F in (4.12), the transfer function can be rewritten as (with C_F neglected for simplicity)

$$R_T = R_o\left(1-g_m R_F\right) \cdot \frac{1+sg_m L_F/\left(g_m R_F-1\right)}{1+sR_oC_t+s^2\left(R_oR_FC_iC_o+L_FC_iR_o/r_o\right)+s^3R_oL_FC_iC_o}, \tag{4.25}$$

where

$$C_t = \left(1+R_F/r_o\right)C_i + C_o. \tag{4.26}$$

In the same way, ζ and m can be calculated as

$$\zeta = \frac{1}{2}\frac{\sqrt{R_o}C_t}{\sqrt{R_FC_iC_o+L_FC_i/r_o}} \tag{4.27}$$

and

$$m = \frac{g_mL_F}{\left(g_mR_F-1\right)R_oC_t}. \tag{4.28}$$

Similarly to the previous case, the introduction of L_F would increase m, but simultaneously decreases ζ. Fortunately, we also have a g_m term, thereby achieving a better controllability over ζ and m. In order to evaluate the impact of L_F on ζ, we define a degradation factor of ζ as

$$r_d \equiv \frac{\zeta|_{L_F=0}}{\zeta} = \sqrt{1+\frac{L_F/r_o}{R_FC_o}}. \tag{4.29}$$

By substituting (4.28) into (4.29), we can obtain the following approximation.

$$r_d \approx m\left(1+\frac{C_i}{C_o}\left(1+R_F/r_o\right)\right)\frac{1}{1+g_mr_o}, \tag{4.30}$$

which implies that small R_F is advantageous for a high BWER.

Now, we apply the aforementioned theories to the design example presented in Figure 4.3. With design parameters of the inverter size and the inductance, we can

calculate the inductance required to achieve a targeted value of *m*. In Figure 4.9, the inductor values needed for $m = 0.5$ are plotted with respect to the inverter size. The required inductance can be as high as 7 nH, which is not realizable on a real silicon. Also, a severe degradation of the damping factor is found, though not shown in this figure. Alternatively, *m* can be lowered and several points are picked for the verification of the theory. With $m = 0.4$, a proper value of the inductance can be calculated for different inverter sizes. As shown in Figure 4.10, by employing an inductive peaking for the inverter-based TIA, an average BWER of 1.85 can be achieved. This result is also compared with the theoretical data, showing that the overall tendency of the theoretical estimation well agrees with that of the simulation results.

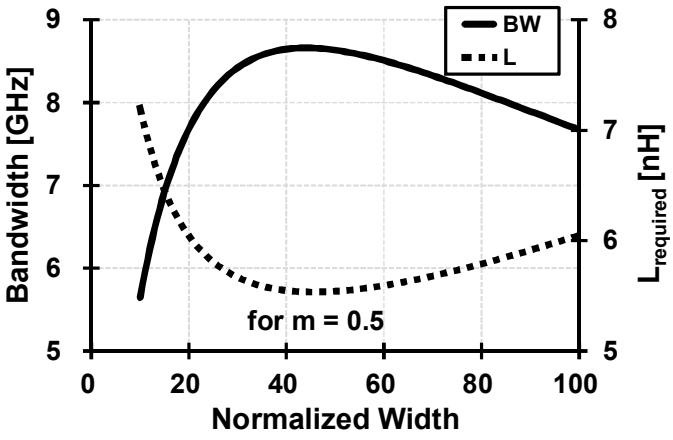


Fig. 4.9 Required inductance to achieve *m* of 0.5.

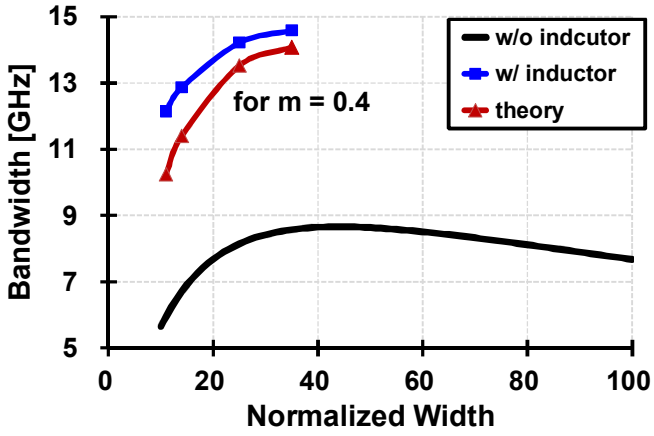


Fig. 4.10 Enhanced bandwidth with *m* of 0.4.

4.4 Circuit Implementation

The block diagram of the implemented optical receiver front-end is shown in Figure 4.11. After the TIA, the single-to-differential (S2D) converter enables a differential signaling. Then, the signal is further amplified by 4-stage limiting amplifiers (LAs) to raise the signal swing to a detectable level. Finally, the signal is driven outside the chip by a 50-Ω driver. For the stabilization of the common-mode level of the differential signal, an offset-cancellation loop is also implemented.

For the bandwidth improvement of the TIA, the resistive and inductive feedback is employed. As discussed in the previous section, a small value of R_F is advantageous in terms of bandwidth extension. In this design, to maximize the bandwidth,

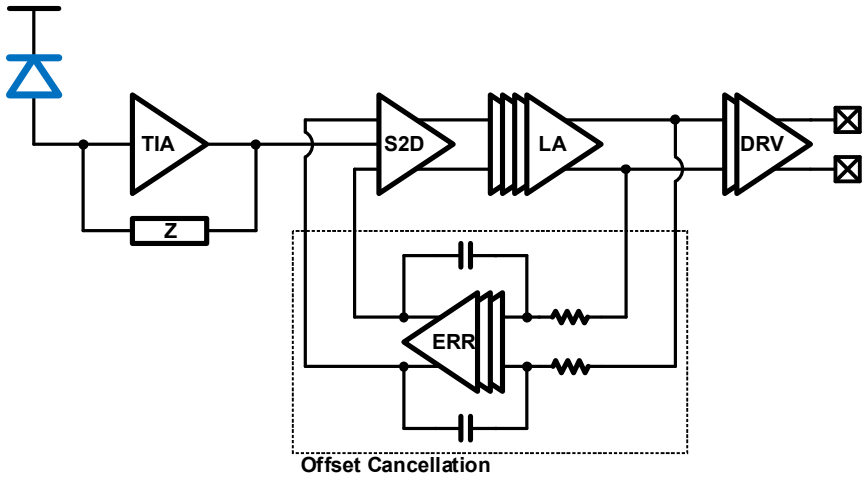
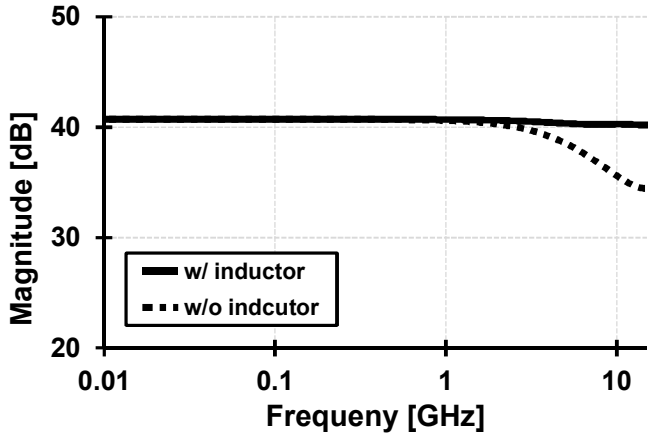


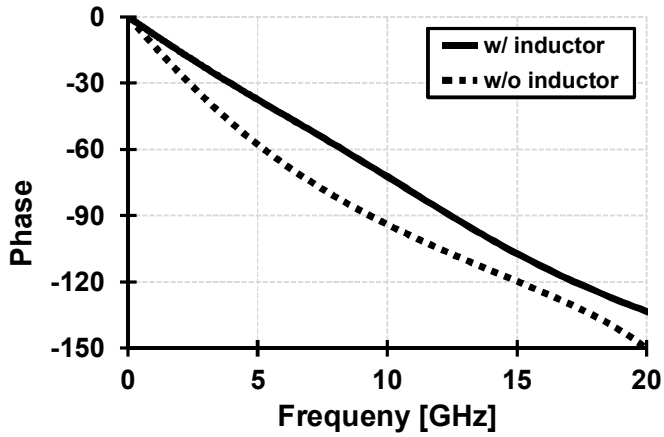
Fig. 4.11 Overall block diagram of optical receiver front-end.

R_F is chosen to be 150 Ω at the cost of the decreased gain, hence the degraded noise performance. The effect of the inductor is simulated and the results are presented in Figure 4.12. The bandwidth can be extended from 11.4 GHz to 25.2 GHz, showing the BWER of 2.2 and the phase response is also remarkably improved as shown in Figure 4.12(b), ensuring a good eye quality at 20 Gb/s. The simulated input-referred noise is 3.9 μA_{rms}, meaning the achievable receiver sensitivity is estimated to be better than -10 dBm [87]. The TIA draws 1.3 mA from a 1.0-V supply, resulting in the power consumption of 1.3 mW.

The implementation of the S2D converter is described in Figure 4.13(a). The S2D consists of an RC low-pass filter and two-cascaded differential amplifiers. The RC time constant is chosen to be sufficiently small to suppress undesired DC wandering. At the first stage, the conventional inductor-shunt peaking is employed for



(a)

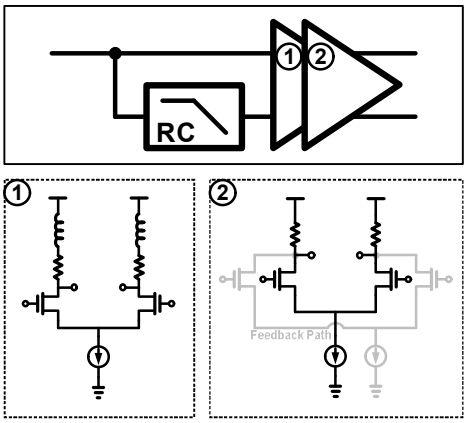


(b)

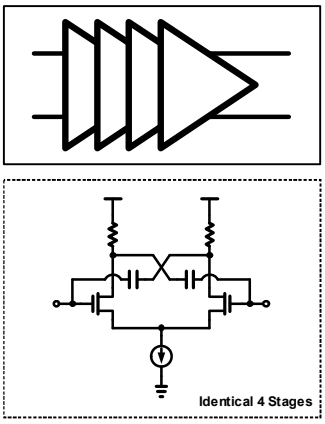
Fig. 4.12 Improvement of TIA frequency response: (a) magnitude response and (b) phase response

bandwidth extension. An additional input path is enabled at the second stage for the

offset-cancellation feedback loop. The S2D converter consumes 8 mW in total. The LA is realized in a similar manner as shown in Figure 4.13(b). It is composed of 4 identical differential amplifiers. Due to a tight restriction on available silicon area, it is impossible to introduce inductors to every LA stage. Instead, a negative Miller



(a)



(b)

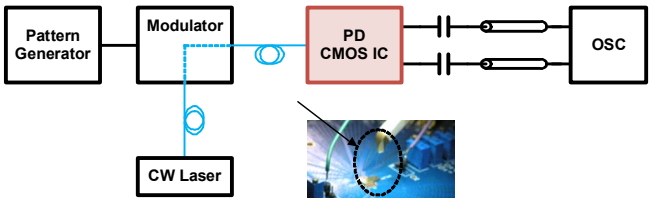
Fig. 4.13 Implementation of (a) S2D converter and (b) LA.

capacitance is employed here, providing a good BWER with much smaller area. The implemented LA consumes 4 mW per stage, hence the total power of 16 mW. The total gain and the bandwidth of the entire front-end (excluding output driver) are found to be, by simulation, 78 dBΩ and 11 GHz, respectively.

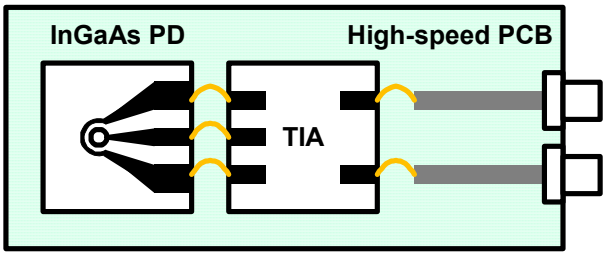
4.5 Measurement Results

The prototype chip is implemented in 65-nm CMOS technology. The overall measurement setup is described in Figure 4.14. For optical measurements, a single-mode fiber is directly coupled onto the InGaAs PD using a passive alignment. The transmitted data is generated from an electrical pattern generator, then converted to an optical signal using an external MZI module as shown in Figure 4.14(a). The wavelength of the CW used in this experiment is 1550 nm. The differential output of the TIA is wire-bonded to the test PCB and then driven to a sampling oscilloscope for eye-diagram measurements as illustrated in Figure 4.14(b). In Figure 4.14(c), the photomicrograph of the prototype chip is presented, showing the hybrid integration of the EPIC using wire bonding.

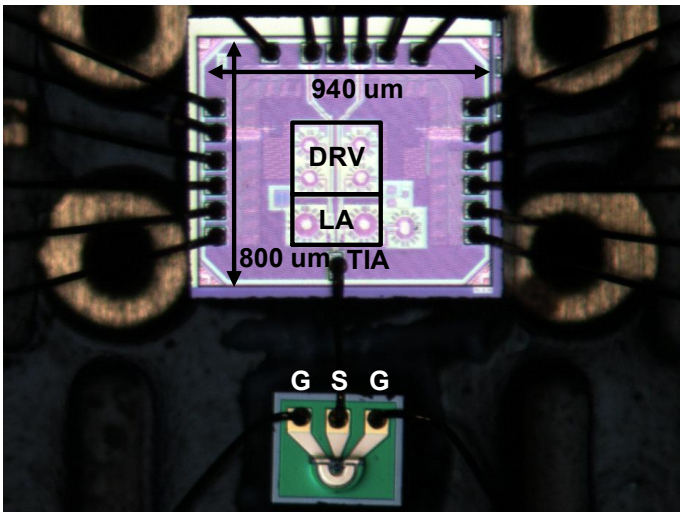
The eye diagrams measured at the data rates of 10 Gb/s and 20 Gb/s are shown in Figure 4.15(a) and (b), respectively. The average input optical power is set to -3 dBm for both cases. For the data rate of 20 Gb/s, the front-end chip consumes a total power of 45.3 mW, which corresponds to an energy efficiency 1.27 pJ/b.



(a)

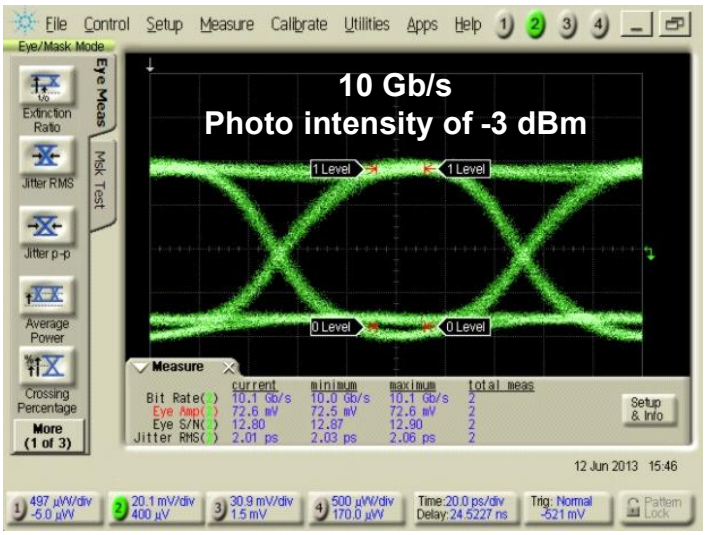


(b)

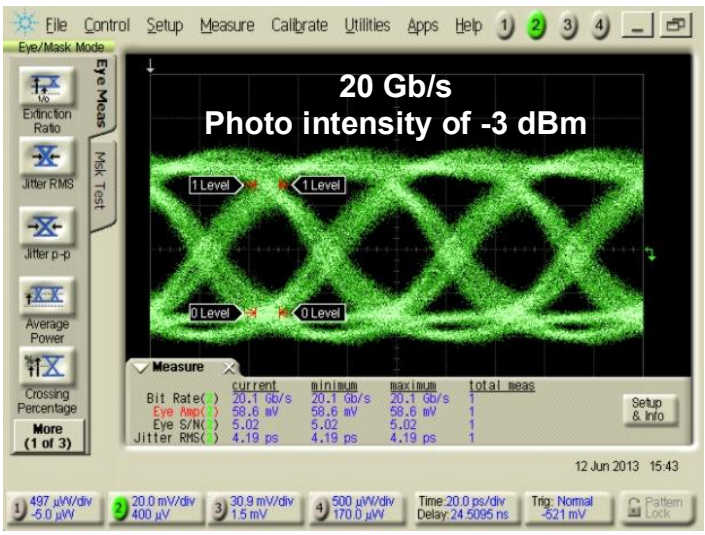


(c)

Fig. 4.14 (a) Description of (a) measurement setups, (b) test board, and (c) chip microphotograph.



(a)



(b)

Fig. 4.15 Measured eye diagrams at the data rate of (a) 10 Gb/s and (b) 20 Gb/s .

Chapter 5

Bandwidth- and Power-Scalable Optical Receiver Front-End

5.1 Overview

For fast delivery of high-quality digital contents, service providers are required to handle ever-increasing data rates. Reflecting this trend, communication standards such Ethernet and InfiniBand demand 40+ Gb/s/lane links for the total throughput of 400 Gb/s and higher. For such high data rates, an optical link has become the most promising solution, gradually replacing the conventional copper-based links. However, there are few 40+ Gb/s/lane optical transceiver implemented without relying on SOI or BiCMOS technologies [54], [95], [96]. Another requirement is the low link power consumption and its scalability is a desirable feature for energy-efficient

communication systems. To meet such demands, a bandwidth- and power-scalable optical receiver front-end operating up to 40 Gb/s is implemented in CMOS technology.

5.2 Bandwidth and Power Scalability

The bandwidth of the optical front-end should be chosen in accordance with the incoming data rate for better noise performance. For this reason, scaling simultaneously both the bandwidth and the power should be a key feature for the optical receiver front-end. In the case of a logic inverter, controlling V_{DD} is the most effective way of bandwidth and power scaling, which is known as dynamic voltage and frequency scaling (DVFS) and widely adopted in electrical links. However, when it comes to an amplifier, changing V_{DD} is less influential in the bandwidth as shown in Figure 5.1 because the input resistance of the amplifier does not scale down.

Meanwhile, in Chapter 4, it was thoroughly investigated that inserting an inductor in series with the feedback resistor has an advantage of bandwidth enhancement. For a better understanding, previously defined parameters, ζ and m are revisited here:

$$\zeta = \frac{1}{2}\frac{\sqrt{R_o}C_t}{\sqrt{R_F C_i C_o + L_F C_i / r_o}} \approx \frac{1}{2}\frac{C_t}{\sqrt{g_m}\sqrt{R_F C_i C_o + L_F C_i / r_o}}. \tag{5.1}$$

$$m = \frac{g_m L_F}{\left(g_m R_F - 1\right) R_o C_t} \approx \frac{g_m L_F}{R_F C_t}. \tag{5.2}$$

It is worth noting that ζ is inversely proportional to the square root of g_m while m is proportional to g_m. Because increasing V_{DD} of the inverter increases g_m without af-

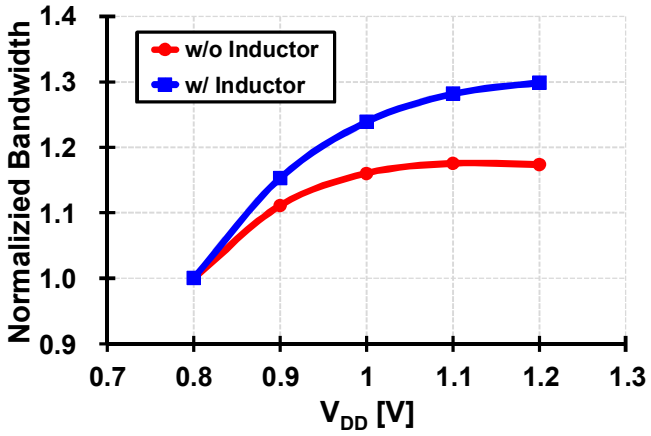


Fig. 5.1 Enhanced bandwidth controllability.

fecting the other parameters, especially the capacitances, the bandwidth enhancement factor can be controlled by adjusting V_{DD}. This is quite significant in that the bandwidth tuning is generally not easily achievable due to difficulties of adjusting the passive elements such as an inductor and a capacitor. By employing the inductive feedback for the inverter-based amplifier, we can achieve both the bandwidth and power scaling, thereby expecting better energy efficiencies over a wide range of data rates.

5.3 G_m Stabilization

Because the performance of the inverter-based amplifier is susceptible to G_m, it is quite important to stabilize G_m against PVT variations. For that purpose, we employ

a G_m regulator that was previously used in [94], to achieve both G_m stabilization and its controllability.

Constant G_m can be achieved by using a replica circuit shown in Fig. 5.2 [94]. The replica circuit is composed of a 2x inverter with source degeneration by a resistor, R_{EXT}, and a 1x inverter. At a low V_{DD}, when the source-degeneration effect is not significant, I_1 is larger than I_2. However, as V_{DD} increases, the source degeneration of the 2x inverter strengthens, making I_2 larger than I_1. For this reason, two I_D–V_{DD} curves intersect at two points as shown in the figure. At point 2, the currents, I_1 and I_2 are given as follows:

$$I_1 = 2\beta_p \left(V_{DD} - V_{G1} - \left|V_{THP}\right|\right)^{\alpha}. \tag{5.3}$$

$$I_1 = 2\beta_n \left(V_{G1} - V_{THN} - I_1 R_{EXT}\right)^{\alpha}. \tag{5.4}$$

$$I_2 = \beta_p \left(V_{DD} - V_{G2} - \left|V_{THP}\right|\right)^{\alpha}. \tag{5.5}$$

$$I_2 = \beta_n \left(V_{G2} - V_{THN}\right)^{\alpha}. \tag{5.6}$$

They are based on the alpha-power model that approximates the current equations rather than sticking to the complicated short-channel equation. The parameters, β_p and β_n in the above equations, contain technology-dependent parameters fitted to actual simulation results. The alpha-power model for the given technology specifies that β_n is twice as large as β_p with the same W/L ratio and α has a value of 1.3. And then, let's assume that

$$I = I_1 = I_2. \tag{5.7}$$

With this new condition, equating (5.3) and (5.5) gives

$$V_{G1} - \frac{V_{G2}}{2^{1/\alpha}} = \left(1 - \frac{1}{2^{1/\alpha}}\right)\left(V_{DD} - V_{THP}\right). \tag{5.8}$$

And equating (5.4) and (5.6) results in

$$I = \frac{1}{R_{EXT}}\left[\left(V_{G1} - \frac{V_{G2}}{2^{1/\alpha}}\right) - \left(1 - \frac{1}{2^{1/\alpha}}\right)V_{THN}\right]. \tag{5.9}$$

By substituting (5.8) into (5.9), I can be rewritten as

$$I = \frac{1}{R_{EXT}}\left(1 - \frac{1}{2^{1/\alpha}}\right)\left(V_{DD} - V_{THP} - V_{THN}\right). \tag{5.10}$$

Meanwhile, using (4.5) and (4.6), following equations can be derived:

$$V_{DD} - V_{G2} - V_{THP} = \frac{\beta_n^{1/\alpha}\left(V_{DD} - V_{THP} - V_{THN}\right)}{\beta_p^{1/\alpha} + \beta_n^{1/\alpha}}. \tag{5.11}$$

$$V_{G2} - V_{THN} = \frac{\beta_p^{1/\alpha}\left(V_{DD} - V_{THP} - V_{THN}\right)}{\beta_p^{1/\alpha} + \beta_n^{1/\alpha}}. \tag{5.12}$$

By definition, G_m of the replica inverter is given by

$$G_m = \frac{\partial I}{\partial\left(V_{DD} - V_{G2}\right)} + \frac{\partial I}{\partial V_{G2}} = \frac{\alpha I}{V_{DD} - V_{G2} - V_{THP}} + \frac{\alpha I}{V_{G2} - V_{THN}}. \tag{5.13}$$

Substituting (5.10), (5.11), and (5.12) into (5.13) results in the final expression for G_m that

$$G_m = \frac{1}{R_{EXT}}\left(\alpha - \frac{\alpha}{2^{1/\alpha}}\right)\left[2 + \left(\frac{\beta_p}{\beta_n}\right)^{1/\alpha} + \left(\frac{\beta_n}{\beta_p}\right)^{1/\alpha}\right]. \tag{5.14}$$

Note that G_m is now dependent only on R_{EXT} and the technology-dependent parameters, while it has nothing to do with process and temperature variations.

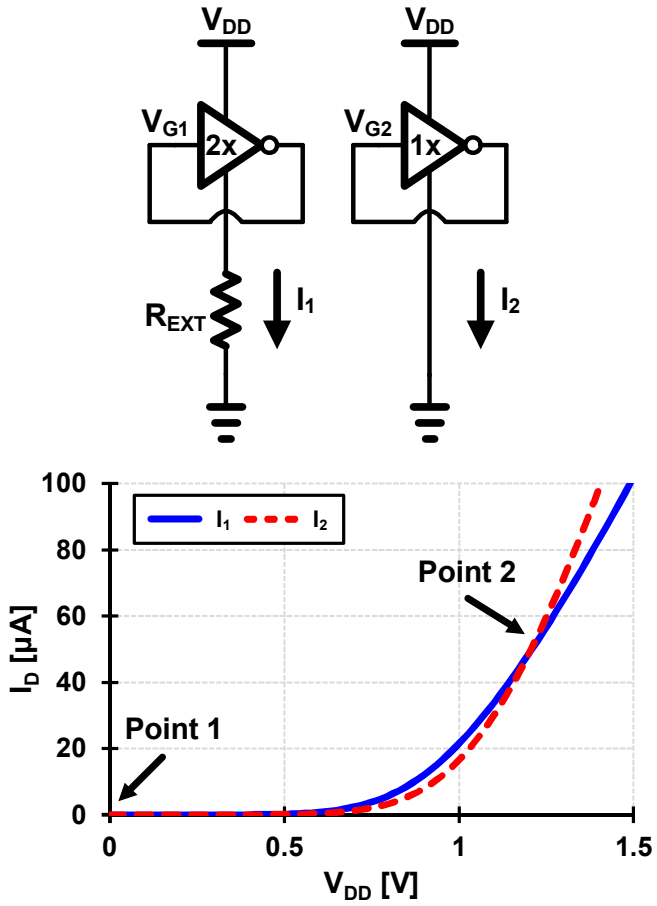


Fig. 5.2 Basic concept of constant-Gm bias generation.

Figure 5.3 details the implementation of a constant-G_m bias generator. To force the currents I_1 and I_2 to be equal, a feedback loop is enabled, and the resulting condition exactly corresponds to the point where (5.14) is derived, which means G_m of the 1x inverter is now made constant with V_2. Non-ideal distribution of the power sup-

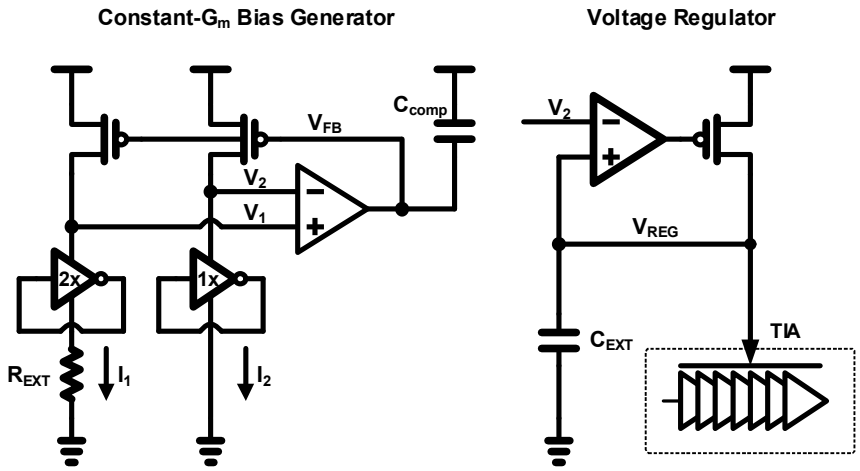
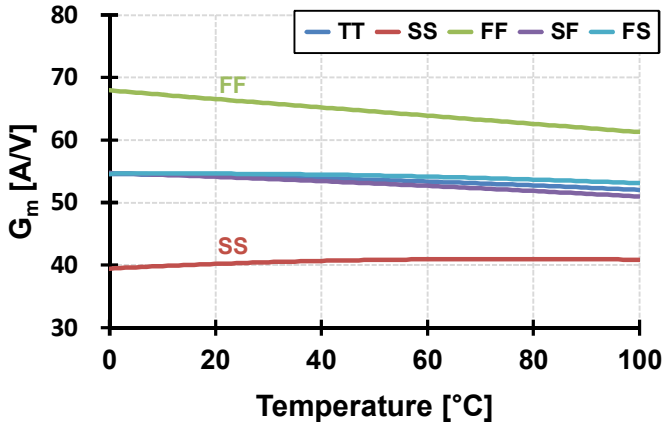


Fig. 5.3 G_m regulator with constant-G_m bias generator and voltage regulator.

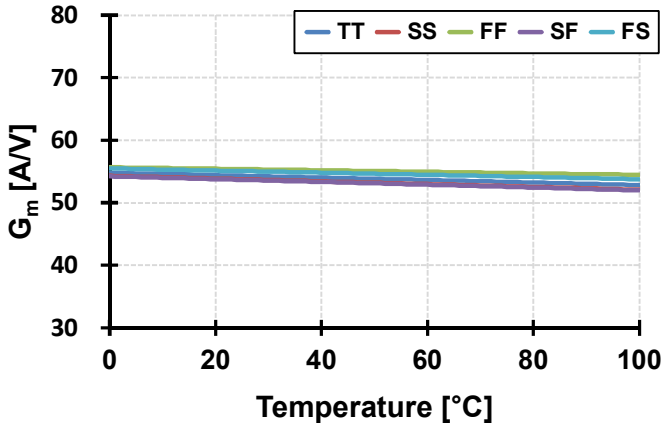
ply, however, mandates the use of a voltage regulator as shown in Figure 5.3, instead of directly using V_{FB} for biasing the PMOS gate. The voltage regulator generates V_{REG} which is exactly the same as V_2, thereby providing constant G_m for the driver which is appropriately scaled to save the power consumption by the replica circuit. Because the multiple feedback loops are enabled in the constant-G_m bias generator, stability issues should be carefully addressed. With the given conditions, it can be found that the feedback loop in the replica circuit is reduced to a three-pole system. Because the two poles at V_1 and V_2 are close to each other, the feedback loop can be made stable by taking another pole at V_{FB} far away from them, forming a dominant pole at the output of the amplifier. This can be easily accomplished by adding a compensating capacitor, C_{comp} as shown in Figure 5.3. Adding C_{comp} brings another advantage that the supply noise on V_2 can be further rejected, resulting in an

enhanced supply noise rejection for the final regulator output, V_{REG}.

With the proposed constant G_m bias generator, against any process and temperature variations, V_{REG} is adequately adjusted to keep G_m of the driver constant, which



(a)



(b)

Fig. 5.4 G_m variations against process and temperature corners (a) w/o G_m regulator and (b) w/ G_m regulator.

is well verified by the simulation results presented in Figure 5.4.

5.4 Overall Block Diagram of Receiver

The block diagram of the implemented optical receiver is shown in Figure 5.5. A high-gain TIA stage consists of seven cascaded inverter amplifiers in a single-ended configuration for the low power consumption. A G_m regulator provides a tunable bandwidth and a regulated supply voltage for the TIA. A DC-offset cancellation loop is employed to ensure the optimum bias condition for the TIA by forcing the average level of the TIA output to be V_{REF} which is generated from the replica TIA. A single-to-differential (S2D converter and an output driver are also implemented to drive a 100-Ω differential transmission line.

Figure 5.6 shows the detailed implementation of the TIA. The inductive feedback

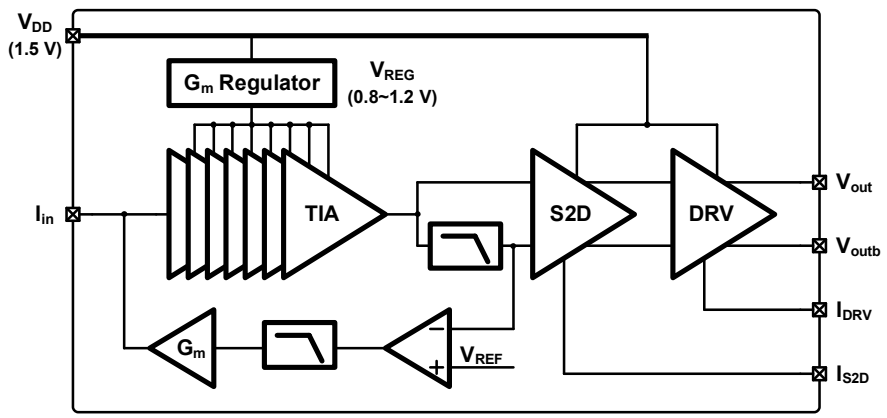


Fig. 5.5 Overall architecture of implemented optical receiver front-end.

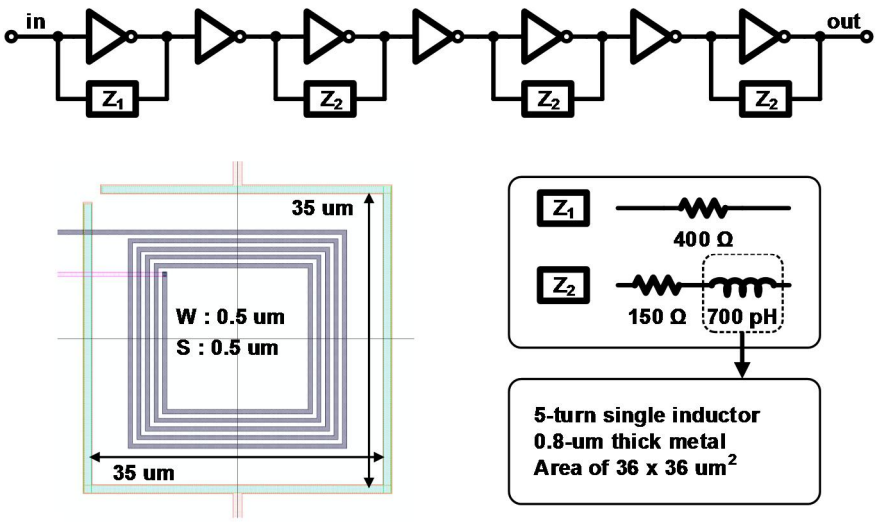


Fig. 5.6 Implementation of TIA.

is not applied to the first stage, because the effect of bandwidth extension is not significant when the feedback resistance is large as discussed previously. For a high-sensitivity function, the gain of the first stage should be sufficient, thus the feedback resistance at the first stage is raised as high as possible. Instead, we employed an inductive feedback for the subsequent stages. The on-chip spiral inductors are compactly designed for the optimum magnitude and phase responses. Because, in this application, the Q factor does not affect the overall performance, the narrow width of the inductor is chosen for a compact size, but we carefully estimated the series resistance of the inductor. The TIA is designed to exhibit a high gain of 10 kΩ, expecting an excellent sensitivity.

With V_{REG} tuned by the G_m regulator, a wide range of the bandwidth (10–20 GHz)

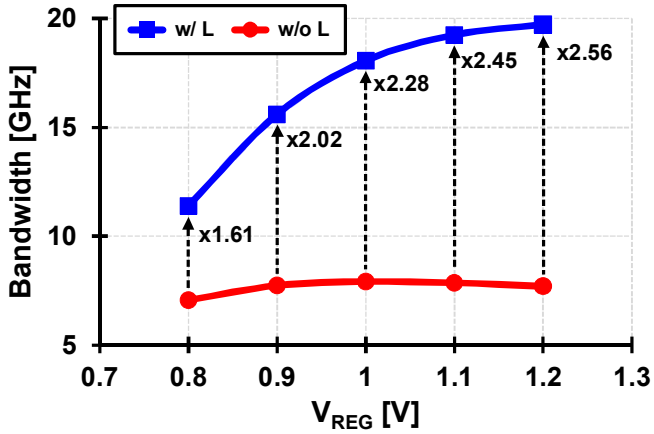


Fig. 5.7 Effect of bandwidth enhancement by introducing inductors.

can achieved, allowing a bandwidth and power scalability. As shown in Figure 5.7, the bandwidth scalability is significantly improved by employing the inductors. Note that the bandwidth is unchanged with varying V_{REG} without the inductors while the bandwidth scalability is greatly improved with the inductors employed.

The S2D and the output driver are implemented with PMOS-based differential pairs to interface with the low common level output of the TIA.

The performance of the front-end circuit is verified by extensive SPICE simulations. For an accurate estimation, we included all the layout parasitic effects and the inductor model extracted from an EM simulator. The supply voltage of the TIA, V_{REG} is appropriately controlled according the data rate. Figure 5.8 shows 10-Gb/s eye diagrams for input conditions of 20, 50, 100, 150 μA_{pp}, respectively with a fixed V_{REG} of 0.9 V. For 28 Gb/s, V_{REG} is raised to 1.0 V and eye diagrams with the same input conditions are shown in Figure 5.9. For 40 Gb/s and 50 Gb/s, we raised

V_{REG} further to 1.2 and 1.3 V, respectively and the simulation results are presented in Figure 5.10 and 5.11. Figure 5.12 summarizes the power consumptions at each data rate and the corresponding energy efficiencies. As expected, a good power scalability can be achieved while the energy efficiency does not scale well with extremely low V_{REG} because of reduced voltage headroom. Noise simulations are also conducted as shown in Figure 5.13. From the simulation results, it is expected that the achievable sensitivities are -15, -11, -8, and -6 dBm at 10, 28, 40, and 50 Gb/s, respectively.

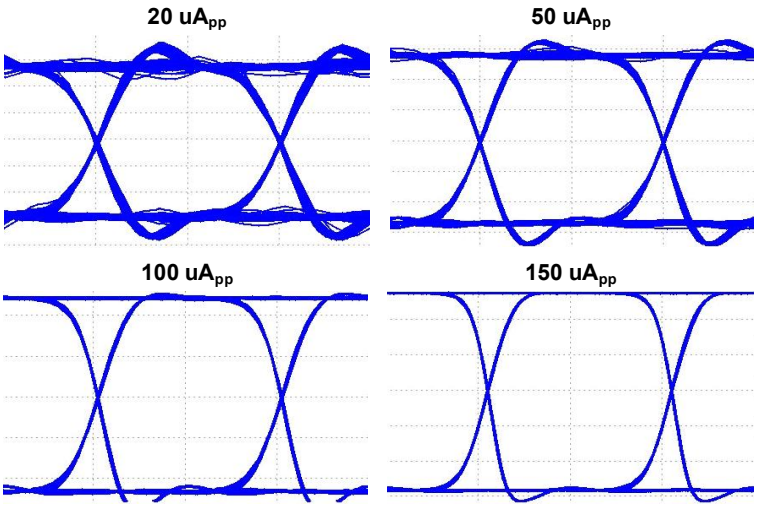


Fig. 5.8 Simulated eye diagrams at 10 Gb/s with V_{REG} of 0.9 V.

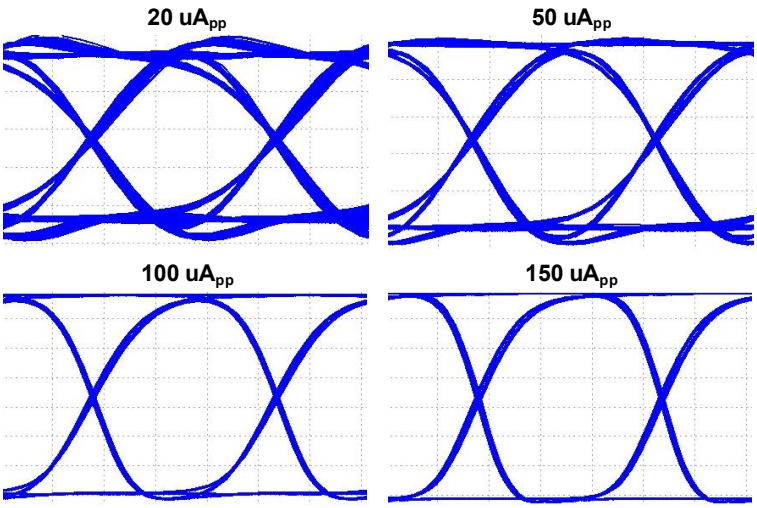


Fig. 5.9 Simulated eye diagrams at 28 Gb/s with V_{REG} of 1.0 V.

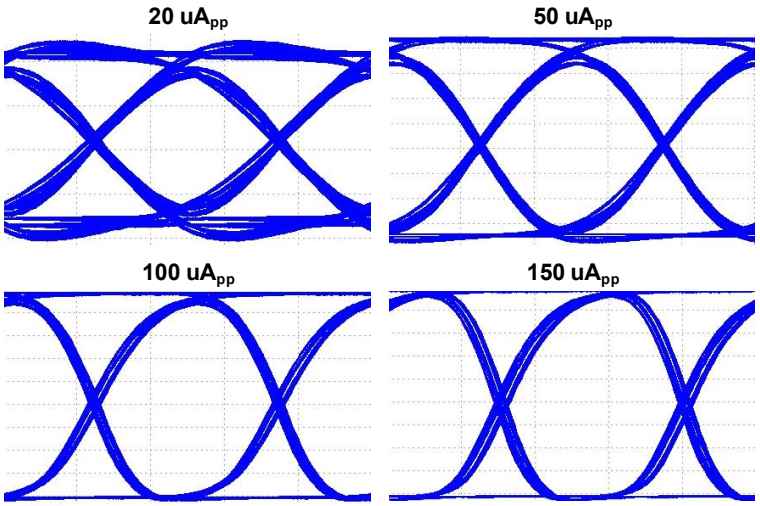


Fig. 5.10 Simulated eye diagrams at 40 Gb/s with V_{REG} of 1.2 V.

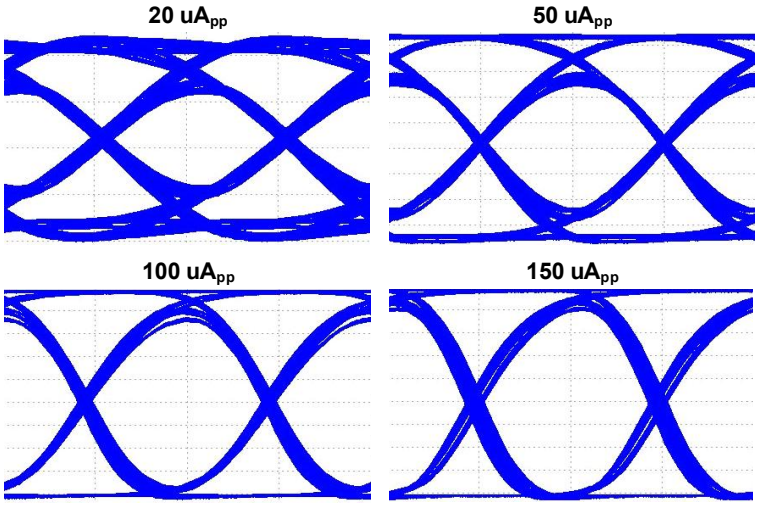
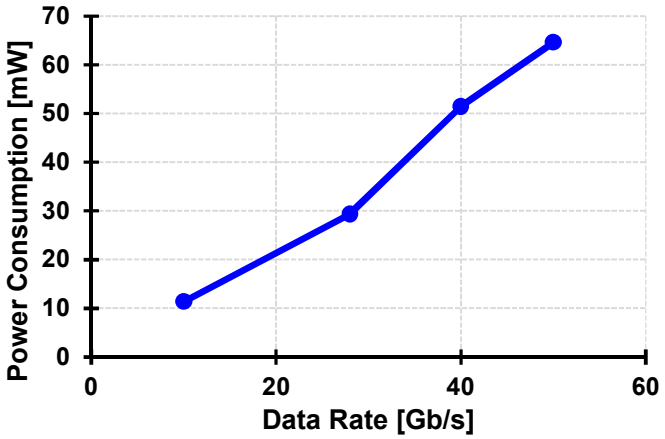
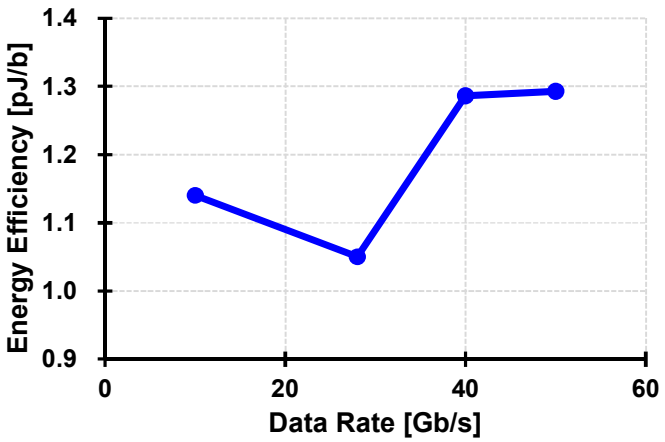


Fig. 5.11 Simulated eye diagrams at 50 Gb/s with V_{REG} of 1.3 V.



(a)



(b)

Fig. 5.12 Simulated (a) power consumption and (b) energy efficiency with respect to data rate.

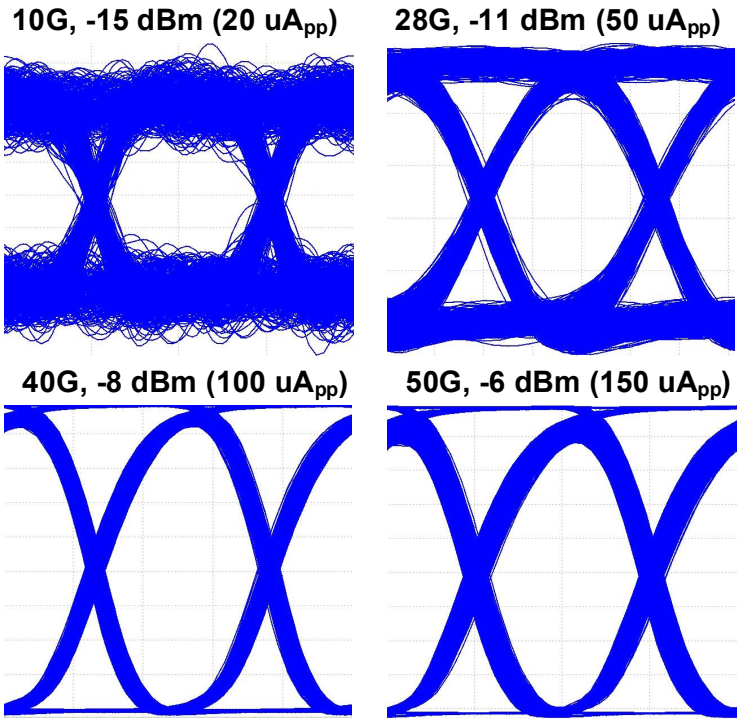


Fig. 5.13 Noise simulation results for various data rates.

5.5 Measurement Results

The optical receiver front-end chip is implemented in 65-nm CMOS technology. The chip microphotograph and measurement setups are shown in Figure 5.14. Electrically and optically (with 28-Gb/s InGaAs PD) measured eye diagrams with 25-Gb/s PRBS data inputs are presented in Figure 5.15. For full characterization, over-

all chip performance is electrically measured with an on-board PD emulation. Bathtub curves measured at 25 and 40 Gb/s are presented in Figure 5.16. As shown in Figure 5.17, the implemented front-end achieves the maximum sensitivities of -12 and -8 dBm at 25 and 40 Gb/s, respectively. The bandwidth and power scalability function is also verified. Bathtub curves measured with V_{REG} of 0.9 and 1.0 V in Figure 5.18 reveal that the excessive bandwidth leads to the degradation of the noise performance, resulting in a narrower opening of the bathtub curve. For the verification of the scalability function, the minimum V_{REG} satisfying respective sensitivity conditions is found for each data rate. As shown in Figure 5.19, both the power consumption and the energy efficiency are well scaled except for the lower data rates and this is because the TIA performance degrades dramatically as V_{REG} gets closer to the minimum turn-on value as mentioned in the previous section. Including the power consumption of the S2D converter (6 mW) and the output driver (6 mW), the entire front-end consumes 32.8 and 51.6 mW at the data rate of 25 and 40 Gb/s, respectively. Table 5.1 summarizes and compares the front-end performances with state-of-the-art works. Owing to the scalability function and the high transimpedance gain, the proposed front-end achieves a constant energy efficiency over a wide range of data rates and excellent sensitivities.

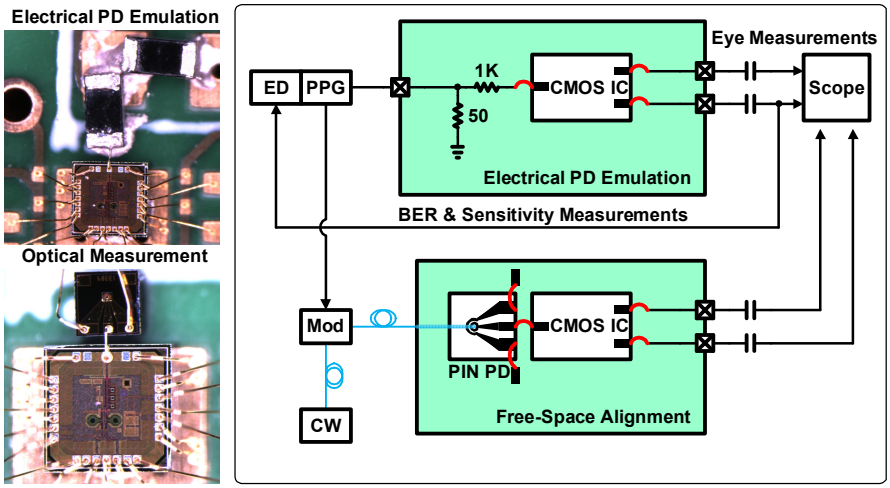
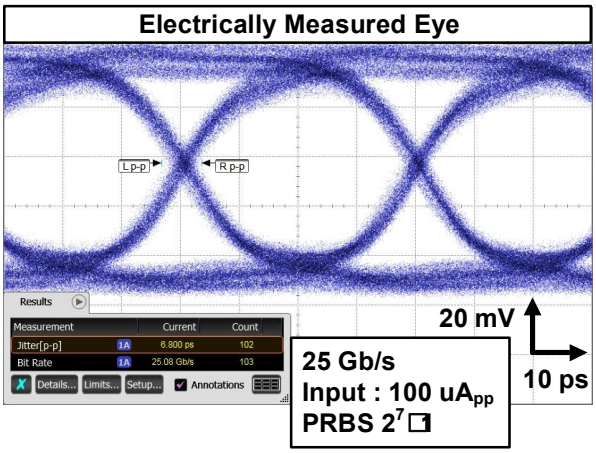
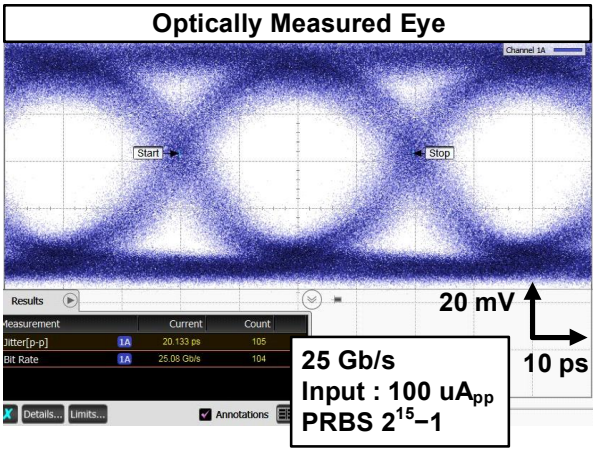


Fig. 5.14 Chip microphotograph and measurement setups.

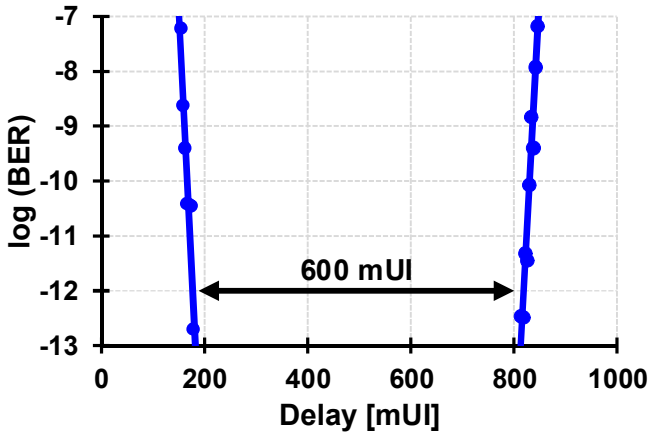


(a)

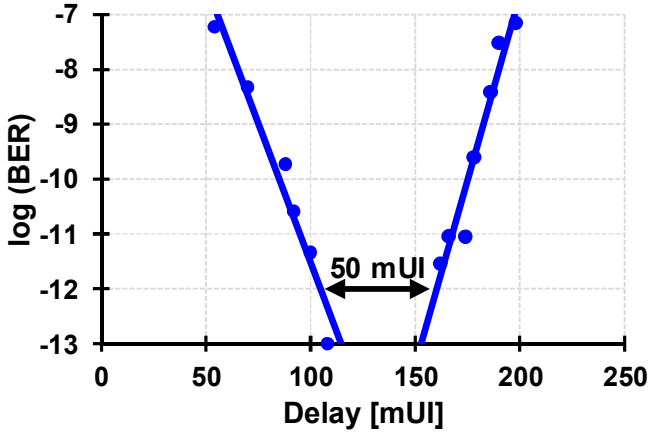


(b)

Fig. 5.15 25-Gb/s eye diagrams measured (a) electrically and (b) optically.



(a)



(b)

Fig. 5.16 Measured bathtub curves at (a) 25 Gb/s and (b) 40 Gb/s.

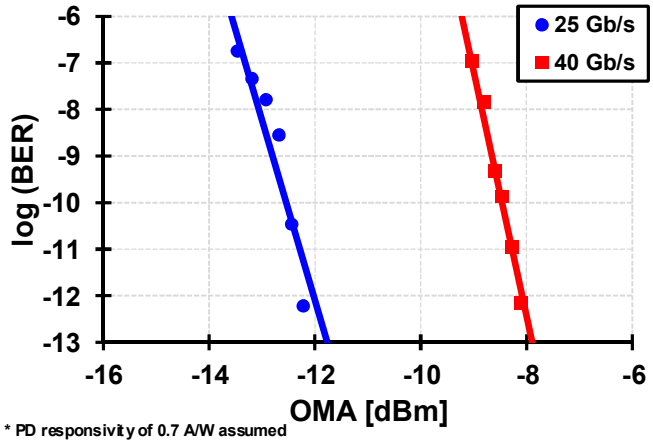


Fig. 5.17 Measured sensitivities at 25 Gb/s and 40 Gb/s.

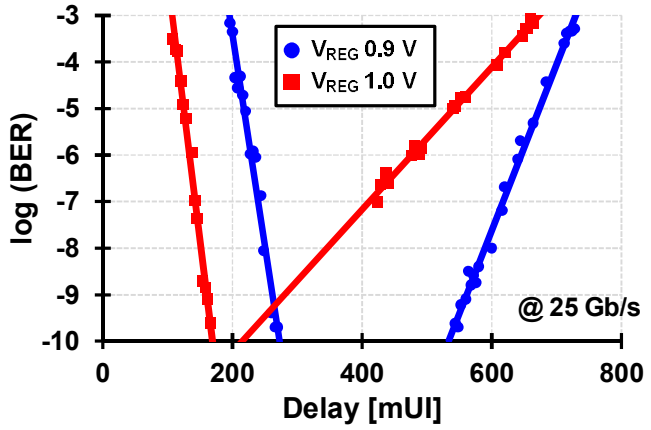


Fig. 5.18 Effectiveness of bandwidth scaling.

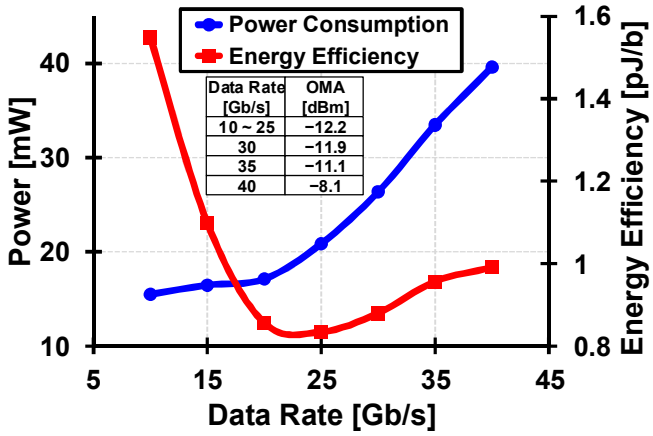


Fig. 5.19 Measured power and energy efficiency scalability.

Table 5.1 Performance comparisons with state-of-the-art works

	VLSI '16 [96]	ISSCC '14 [71]	ISSCC '15 [41]	JSSC '15 [75]	This Work	
Data Rate [Gb/s]	56	28	25	26.5	25	40
Power [mW]	419	28.8	50	35.7	32.8	51.6
Energy Efficiency [pJ/b]	7.5	1.03	2	1.35	1.31	1.29
Sensitivity [uA_{pp}]	220	63	127	184	42	108
Technology	180 nm BiCMOS	28 nm CMOS	28 nm CMOS	65 nm CMOS	65 nm CMOS	

Chapter 6

Conclusion

With enormously increasing demands for a total aggregate data bandwidth, the capability of the legacy electrical link has reached its physical limit. Alternatively, the optical link can provide much higher data bandwidth, and hence, is regarded as a promising solution that can replace the electrical link. In this thesis, we summarize silicon photonic technologies in various aspects. Not only the current status of silicon photonics but also remaining challenges are discussed thoroughly. From these observations, the requirements for the optical interface circuits are drawn and we attempt to fully reflect them in the receiver circuit design. For a successful replacement of electrical links with optical links, we have to focus on the following requirements: high speed, low power, and CMOS implementation. The most appropriate topology satisfying those requirements is an inverter-based TIA. However, the inverter-based TIA is also bandwidth-limited and a proper bandwidth extension technique is essential to future high-speed applications. In this regard, this thesis is

devoted to investigating operating principles and presenting design methodologies for the proposed optical receiver front-end with a novel bandwidth extension technique. This work is significant in that it first provides both a qualitative and a quantitative analyses of the inverter-based TIA with resistive and inductive feedback. Based on this topology, the optical receiver front-ends operating at 20 and 40 Gb/s, respectively, with low power consumption are successfully demonstrated. Especially, the 40-Gb/s receiver front-end features a bandwidth- and power-scalability function which is necessary for energy-efficient link systems.

Bibliography

[1] Cisco Systems, Inc. [Online]. Available: http://www.cisco.com

[2] R. A. Soref, "Silicon-based optoelectronics," Proc. IEEE, vol. 81, No. 12, pp. 1687–1706, Dec. 1993.

[3] N. Izhaky et al., and D. J. Friedman, "Development of CMOS-compatible integrated silicon photonics devices," IEEE J. Selected Topics in Quantum Electronics, vol. 12, no. 6, pp. 1688–1698, Nov./Dec. 2006.

[4] Y. A. Vlasov, "Silicon CMOS-integrated nano-photonics for computer and data communication beyond 100G," IEEE Communication Magazine, vol. 50, No. 2, pp. s67–s72, Feb. 2012.

[5] C. Gunn, "CMOS photonics for high-speed interconnects," IEEE Micro, vol. 26, No. 2, pp. 58–66, Mar.–Apr. 2006.

[6] F. Y. Liu et al., "10-Gb/s 5.3-mW optical transmitter and receiver circuits in 40-nm CMOS," IEEE J. Solid-State Circuits, vol. 47, no. 9, pp. 2049–2067, Sep. 2012.

[7] Y. Arakawa, T. Nakamura, Y. Urino, and T. Fujita, "Silicon photonics for next generation system integration platform," IEEE Communication Magazine, vol. 51, No. 3, pp. 72–77, Mar. 2013.

[8] B. Zhang et al., "A 28Gb/s multi-standard serial-link transceiver for backplane applications in 28nm CMOS," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2015, pp. 1–3.

[9] T. Kawamoto et al., "Multi-standard 185fs_{rms} 0.3-to-28Gb/s 40dB backplane signal conditioner with adaptive pattern-match 36-tap DFE and data-rate-adjustment PLL in 28nm CMOS," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2015, pp. 1–3.

[10] T. Norimatsu et al., “A 25Gb/s multistandard serial link transceiver for 50dB-loss copper cable in 28nm CMOS,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2016, pp. 60–61.

[11] P.-J. Peng, J.-F. Li, L.-Y. Chen, and J. Lee, “A 56Gb/s PAM-4/NRZ transceiver in 40nm CMOS,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2016, pp. 110–111.

[12] J. Im et al., “A 40-to-56Gb/s PAM-4 receiver with 10-tap direct decision-feedback equalization in 16nm FinFET,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2017, pp. 114–115.

[13] G. Steffan et al., “A 64Gb/s PAM-4 transmitter with 4-tap FFE and 2.26pJ/b energy efficiency in 28nm CMOS FDSOI,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2017, pp. 116–117.

[14] T. O. Dickson, H. A. Ainspan, M. Meghelli, “A 1.8pJ/b 56Gb/s PAM-4 transmitter with fractionally spaced FFE in 14nm CMOS,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2017, pp. 118–119.

[15] IEEE P802.3ae 10Gb/s Ethernet Task Force. [Online]. Available: http://www.ieee802.org/3/ae

[16] IEEE P802.3ba 100Gb/s Ethernet Task Force. [Online]. Available: http://www.ieee802.org/3/ba

[17] 100G CLR4 Alliance. [Online]. Available: http://www.clr4-alliance.org

[18] CWDM4 MSA. [Online]. Available: http://www.cwdm4-msa.org

[19] B. Jalali, M. Paniccia, and G. Reed, “Silicon photonics,” IEEE Microwave Magazine, vol. 7, No. 3, pp. 58–68, Jun. 2006.

[20] R. Soref, “The past, present, and future of silicon photonics,” IEEE J. Selected Topics in Quantum Electronics, vol. 12, no. 6, pp. 1678–1687, Nov.–Dec. 2006.

[21] M. Hochberg et al., “Silicon photonics: the next fabless semiconductor

industry," IEEE Solid-States Circuits Magazine, vol. 5, no. 1, pp. 48–58, Winter 2013.

[22] L. Vivien et al., "Zero-bias 40Gbit/s germanium waveguide photodetector on silicon," Optics Express, vol. 20, no. 2, pp. 1096–1101, Jan. 2012.

[23] S. Assefa, F. Xia, and Y. A. Vlasov, "Reinventing germanium avalanche photodetector for nanophotonic on-chip optical interconnects," Nature, vol. 464, no. 4, pp. 80–84, Mar. 2010.

[24] D. J. Thomson et al., "50-Gb/s silicon optical modulator," IEEE Photonics Technology Letter, vol. 24, no. 4, pp. 234–236, Feb, 2012.

[25] E. Temporiti, G. Minoia, M. Repossi, D. Baldi, A. Ghilioni, F. Svelto, "A 56Gb/s 300mW silicon-photonics transmitter in 3D-integrated PIC25G and 55nm BiCMOS technologies," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2016, pp. 404–405.

[26] Q. Xu, B. Schmidt, S. Pradhan, and M. Lipson, "Micrometre-scale silicon electro-optic modulator," Nature, vol. 435, no. 19, pp. 325–327, Mar. 2005.

[27] G. Li et al., "25Gb/s 1V-driving CMOS ring modulator with integrated thermal tuning," Optics Express, vol. 19, no. 21, pp. 20435–20443, Oct. 2011.

[28] M. Rakowski et al., "A 4x20Gb/s WDM ring-based hybrid CMOS silicon photonics transceiver," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2015, pp. 1–3.

[29] B. Analui, D. Guckenberger, D. Kucharski, and A. Narasimha, "A fully integrated 20-Gb/s optoelectronic transceiver implemented in a standard 0.13-μm CMOS SOI technology," IEEE J. Solid-State Circuits, vol. 41, no. 12, pp. 2945–2955, Dec. 2006.

[30] A. Narasimha et al., "A fully integrated 4x 10-Gb/s DWDM optoelectronic transceiver implemented in a standard 0.13 μm CMOS SOI technolo-

gy," IEEE J. Solid-State Circuits, vol. 42, no. 12, pp. 2736–2744, Dec. 2007.

[31] M. Georgas, J. Orcutt, R. J. Ram, V. Stojanovic, "A monolithically-integrated optical receiver in standard 45-nm SOI," IEEE J. Solid-State Circuits, vol. 47, no. 7, pp. 1693–1702, Jul. 2012.

[32] J. F. Buckwalter, X. Zheng, G. Li, K. Raj, and A. V. Krishnamoorthy, "A monolithic 25-Gb/s transceiver with photonic ring modulators and Ge detectors in a 130-nm CMOS SOI process," IEEE J. Solid-State Circuits, vol. 47, no. 6, pp. 1309–1322, Jun. 2012.

[33] C. Sun et al., "Single-chip microprocessor that communicates directly using light," Nature, vol. 528, no. 24, pp. 534–538, Dec. 2015.

[34] R. Meade et al., "Integration of silicon photonics in bulk CMOS," in Symp. VLSI Technology Dig. Tech. Papers, 2014, pp. 228-229.

[35] C. Sun et al., "A monolithically-integrated chip-to-chip optical link in bulk CMOS," IEEE J. Solid-State Circuits, vol. 50, no. 4, pp. 828–844, Apr. 2015.

[36] H. Byun et al., "Bulk-Si photonics technology for DRAM interface," Photonics Research, vol. 2, no. 3, pp. A25–A33, Apr. 2014.

[37] J. Proesel, C. Schow, and A. Rylyakov, "25Gb/s 3.6pJ/b and 15Gb/s 1.37pJ/b VCSEL-based optical links in 90nm CMOS," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2012, pp. 418–420.

[38] A. Cevrero et al., "A 64Gb/s 1.4pJ/b NRZ optical-receiver data-path in 14nm CMOS FinFET," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2017, pp. 482–483.

[39] T. Takemoto, H. Yamashita, T. Yazaki, N. Chujo, Y. Lee, and Y. Matsuoka, "A 4x 25-to-28Gb/s 4.9mW/Gb/s -9.7dBm high-sensitivity optical receiver based on 65nm CMOS for board-to-board interconnects," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2013, pp. 118–119.

[40] H. Morita et al., “A 12x5 two-dimensional optical I/O array for 600Gb/s chip-to-chip interconnect in 65nm CMOS,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2014, pp. 140–141.

[41] Y. Chen et al., “A 25Gb/s hybrid integrated silicon photonic transceiver in 28nm CMOS and SOI,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2015, pp. 1–3.

[42] M. Cignoli et al., “A 1310nm 3D-integrated silicon photonics Mach-Zehnder-based transmitter with 275mW multistage CMOS driver achieving 6dB extinction ratio at 25Gb/s,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2015, pp. 1–3.

[43] M. W. Geis et al., “CMOS-compatible all-Si high-speed waveguide photodiodes with high responsivity in near-infrared communication band,” IEEE Photonics Technology Letters, vol. 19, no. 3, pp. 152–154, Feb. 2007.

[44] M. W. Geis, S. J. Spector, M. E. Grein, J. U. Yoon, D. M. Lennon, and T. M. Lyszczarz., “Silicon waveguide infrared photodiodes with >35 GHz bandwidth and phototransistors with 50 AW^{-1} response,” Optics Express, vol. 17, no. 7, pp. 5193–5204, Mar. 2009.

[45] J. K. Doylend, P. E. Jessop, and A. P. Kights, “Silicon photonic resonator-enhanced defect-mediated photodiode for sub-bandgap detection,” Optics Express, vol. 18, no. 14, pp. 14671–14678, Jun. 2010.

[46] K. Preston, Y. H. D. Lee, M. Zhang, and M. Lipson, “Waveguide-integrated telecom-wavelength photodiode in deposited silicon,” Optics Letters, vol. 36, no. 1, pp. 52–54, Jun. 2011.

[47] K. K. Mehta et al., “Polycrystalline silicon ring resonator photodiodes in a bulk complementary metal-oxide-semiconductor process,” Optics Letters, vol. 39, no. 4, pp. 1061–1064, Feb. 2014.

[48] G. T. Reed, Silicon Photonics: the state of the art, John Wiley & Sons, 2008.

[49] S. B. Samavedam, M. T. Currie, T. A. Langdo, and E. A. Fitzgerald, "High-quality germanium photodiodes integrated on silicon substrates using optimized relaxed graded buffers," Applied Physics Letters, vol. 73, no. 15, pp. 2125–2127, 1998.

[50] L. Colace et al., "Metal-semiconductor-metal near-infrared light detector based on epitaxial Ge/Si," Applied Physics Letters, vol. 72, no. 24, pp. 3175–3177, 1998.

[51] H.-C. Luan et al., "High-quality Ge epilayers on Si with low threading-dislocation densities," Applied Physics Letters, vol. 75, no. 19, pp. 2909–2911, 1999.

[52] S. Fama, L. Colace, G. Masini, G. Assanto, and H.-C. Luan, "High performance germanium-on-silicon detectors for optical communications," Applied Physics Letters, vol. 81, no. 4, pp. 586–588, 2002.

[53] H.-Y. Yu et al., "High-efficiency pin photodetectors on selective-area-grown Ge for monolithic integration," IEEE Electron Device Letters, vol. 30, no. 11, pp. 1161-1163, Nov. 2009.

[54] J. Kim, and J. F. Buckwalter, "A 40-Gb/s optical transceiver front-end in 45 nm SOI CMOS," IEEE J. Solid-State Circuits, vol. 47, no. 3, pp. 615–626, Mar. 2012.

[55] J. Proesel, A. Rylyakov, and C. Schow, "Optical receivers using DFE-IIR equalization," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2013, pp. 130–131.

[56] S. M. Park, and C. Toumazou, "A packaged low-noise high-speed regulated cascade transimpedance amplifier using a 0.6μm N-well CMOS technology," in Proc. European Solid-State Circuits Conf., 2000, pp. 431–434.

[57] S. M. Park, and H.-J. Yoo, "1.25-Gb/s regulated cascade CMOS transimpedance amplifier for gigabit Ethernet applications," IEEE J. Solid-State Circuits, vol. 39, no. 1, pp. 112–121, Jan. 2004.

[58] S. M. Park, J. Lee, and H.-J. Yoo, "1-Gb/s 80-dBΩ fully differential CMOS transimpedance amplifier in multichip on oxide technology for optical interconnects," IEEE J. Solid-State Circuits, vol. 39, no. 6, pp. 971–974, Jun. 2004.

[59] T. Takemoto et al., "A 25-Gb/s 2.2-W 65-nm CMOS optical transceiver using a power-supply-variation-tolerant analog front end and data-format conversion," IEEE J. Solid-State Circuits, vol. 49, no. 2, pp. 471–485, Feb. 2014.

[60] T. Takemoto, H. Yamashita, T. Yazaki, N. Chujo, Y. Lee, and Y. Matsuoka, "A 25-to-28 Gb/s high-sensitivity (–9.7 dBm) 65 nm CMOS optical receiver for board-to-board interconnects," IEEE J. Solid-State Circuits, vol. 49, no. 10, pp. 2259–2276, Oct. 2014.

[61] C. Kromer et al., "A low-power 20-GHz 52-dBΩ transimpedance amplifier in 80-nm CMOS," IEEE J. Solid-State Circuits, vol. 39, no. 6, pp. 885–894, Jun. 2004.

[62] A. A. Abidi, "Gigahertz transimpedance amplifiers in fine line NMOS," IEEE J. Solid-State Circuits, vol. 19, no. 6, pp. 986–994, Dec. 1984.

[63] P.-C. Chiang, J.-Y. Jiang, H.-W. Hung, C.-Y. Wu, G.-S. Chen, and J. Lee, "4x25 Gb/s transceiver with optical front-end for 100 GbE system in 65 nm CMOS technology," IEEE J. Solid-State Circuits, vol. 50, no. 2, pp. 573–585, Feb. 2015.

[64] P. J-W. Lim, A. Y. C. Tzeng, H. L. Chuang, S. A. St. Onge, "A 3.3-V monolithic photodetector/CMOS preamplifier for 531Mb/s optical data link applications," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 1993, pp. 96–97.

[65] T. K. Woodward, A. V. Krishnamoorthy, R. G. Rozier, and A. L. Lentine, "Low-power, small-footprint gigabit Ethernet-compatible optical receiver circuit in 0.25μm CMOS," Electronics Letters, vol. 36, no. 17, pp. 1489–1491, Aug. 2000..

[66] B. Analui, and A. Hajimiri, “Bandwidth enhancement for transimpedance amplifiers,” IEEE J. Solid-State Circuits, vol. 39, no. 8, pp. 1263–1270, Aug. 2004.

[67] M. Ingels, G. Van Der Plas, J. Crols, and M. Steyaert, “A CMOS 18 THzΩ 240 Mb/s transimpedance amplifier and 155 Mb/s LED-driver for low cost optical fiber links,” IEEE J. Solid-State Circuits, vol. 29, no. 12, pp. 1552–1559, Dec. 1994.

[68] T. Yoon, and B. Jalali, “1.25 Gb/s CMOS differential transimpedance amplifier for gigabit networks,” in Proc. European Solid-State Circuits Conf., 1997, pp. 140–143

[69] F. Tavernier, M. S. J. Steyaert, “High-speed optical receivers with integrated photodiode in 130 nm CMOS,” IEEE J. Solid-State Circuits, vol. 44, no. 10, pp. 2856–2867, Oct. 2009.

[70] T. K. Woodward and A. V. Krishnamoorthy, “1 Gbit/s CMOS photoreceiver with integrated detector operating at 850 nm,” Electronics Letters, vol. 34, no. 12, pp. 1252–1253, Jun. 1998.

[71] T.-C. Huang et al., “A 28Gb/s 1pJ/b shared-inductor optical receiver with 56% chip-area reduction in 28nm CMOS,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2014, pp. 144–145.

[72] C. Li et al., “Silicon photonics transceiver circuits with microring resonator bias-based wavelength stabilization in 65 nm CMOS,” IEEE J. Solid-State Circuits, vol. 49, no. 6, pp. 1419–1436, Jun. 2014.

[73] D. Li et al., “A low-noise design technique for high-speed CMOS optical receivers,” IEEE J. Solid-State Circuits, vol. 49, no. 6, pp. 1437–1447, Jun. 2014.

[74] A. Rylyakov et al., “A 25Gb/s burst-mode receiver for rapidly reconfigurable optical networks,” in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2015, pp. 1–3.

[75] S.-H. Chu et al., “A 22 to 26.5 Gb/s optical receiver with all-digital clock and data recovery in a 65 nm CMOS process,” IEEE J. Solid-State Circuits, vol. 50, no. 11, pp. 2603–2612, Nov. 2015.

[76] C. Sun et al., “A 45 nm CMOS-SOI monolithic photonics platform with bit-statistics-based resonant microring thermal tuning,” IEEE J. Solid-State Circuits, vol. 51, no. 4, pp. 893–907, Apr. 2016.

[77] K. Yu et al., “A 25 Gb/s hybrid-integrated silicon photonic source-synchronous receiver with microring wavelength stabilization,” IEEE J. Solid-State Circuits, vol. 51, no. 9, pp. 2129–2141, Sep. 2016.

[78] A. Emami, D. Liu, G. Keeler, N. Helman, and M. Horowitz, “A 1.6Gb/s, 3 mW CMOS receiver for optical communication,” in Symp. VLSI Circuits Dig. Tech. Papers, 2002, pp. 84-87.

[79] S. Palermo, A. Emami, and M. Horowitz, “A 90 nm CMOS 16 Gb/s transceiver for optical interconnects,” IEEE J. Solid-State Circuits, vol. 43, no. 5, pp. 1235–1246, May. 2008.

[80] M. H. Nazari, and A. Emami, “A 24-Gb/s double-sampling receiver for ultra-low-power optical communication,” IEEE J. Solid-State Circuits, vol. 48, no. 2, pp. 344–357, Feb. 2013.

[81] S. Galal, and B. Razavi, “40-Gb/s amplifier and ESD protection circuit in 0.18-µm CMOS technology,” IEEE J. Solid-State Circuits, vol. 39, no. 12, pp. 2389–2396, Dec. 2004.

[82] S. Shekhar, J. S. Walling, and D. J. Allstot, “Bandwidth extension techniques for CMOS amplifiers,” IEEE J. Solid-State Circuits, vol. 41, no. 11, pp. 2424–2439, Nov. 2006.

[83] J. Kim, J.-K. Kim, B.-J. Lee, and D.-K. Jeong, “Design optimization of on-chip inductive peaking structures for 0.13-µm CMOS 40-Gb/s transmitter circuits,” IEEE Trans. Circuits and Systems I, Regular Papers, vol. 56, no. 12, pp. 2544–2555, Dec. 2009.

[84] S. S. Mohan, M. M. Hershenson, S. P. Boyd, and T. H. Lee, "Bandwidth extension in CMOS with optimized on-chip inductors," IEEE J. Solid-State Circuits, vol. 35, no. 3, pp. 346–355, Mar. 2000.

[85] J. Kim, and J. F. Buckwalter, "Bandwidth enhancement with low group-delay variation for a 40-Gb/s transimpedance amplifier," IEEE Trans. Circuits and Systems I, Regular Papers, vol. 57, no. 8, pp. 1964–1972, Aug. 2010.

[86] J. Han et al., "A 20-Gb/s transformer-based current-mode optical receiver in 0.13-μm CMOS," IEEE Trans. Circuits and Systems II, Express Briefs, vol. 57, no. 5, pp. 348–352, May. 2010.

[87] G.-S. Jeong, H. Chi, K. Kim, and D.-K. Jeong, "A 20-Gb/s 1.27pJ/b low-power optical receiver front-end in 65nm CMOS," in proc. IEEE International Symposium on Circuits and Systems, 2014, pp. 1492–1495.

[88] W. Bae, G.-S. Jeong, Y. Kim, H.-K. Chi, and D.-K. Jeong, "Design of silico photonic interconnect ICs in 65-nm CMOS technology," IEEE Trans. Very Large Scale Integration Systems, vol. 24, no. 6, pp. 2234–2243, Jun. 2016.

[89] T. S.-C. Kao, F. A. Musa, A. C. Carusone, "A 5-Gbit/s CMOS optical receiver with integrated spatially modulated light detector and equalization," IEEE Trans. Circuits and Systems I, Regular Papers, vol. 57, no. 11, pp. 2844–2857, Nov. 2010.

[90] A. Sharif-Bakhtiar, and A. C. Carusone, "A 20 Gb/s CMOS optical receiver with limited-bandwidth front end and local feedback IIR-DFE," IEEE J. Solid-State Circuits, vol. 51, no. 11, pp. 2679–2689, Nov. 2016.

[91] W. Bae, G.-S. Jeong, K. Park, S.-Y. Cho, Y. Kim, and D.-K. Jeong, "A 0.36 pJ/bit, 0.025 mm^2, 12.5 Gb/s forwarded-clock receiver with a stuck-free delay-locked loop and a half-bit delay line in 65-nm CMOS technology," IEEE Trans. Circuits and Systems I, Regular Papers, vol. 63, no. 9, pp. 1393–1403, Sep. 2016.

[92] S. T. Flannagan, "Synchronization reliability in CMOS technology," IEEE J. Solid-State Circuits, vol. 20, no. 4, pp. 880–882, Aug. 1985.

[93] T. Sakurai, "Optimization of CMOS arbiter and synchronizer circuits with submicrometer MOSFET's," IEEE J. Solid-State Circuits, vol. 23, no. 4, pp. 901–906, Aug. 1988.

[94] G.-S. Jeong et al., "A 20 Gb/s 0.4 pJ/b energy-efficient transmitter driver utilizing constant-G_m bias," IEEE J. Solid-State Circuits, vol. 51, no. 10, pp. 2312–2327, Oct. 2016.

[95] A. Awny et al., "A dual 64Gbaud 10kΩ 5% THD linear differential transimpedance amplifier with automatic gain control in 0.13μm BiCMOS technology for optical fiber coherent receivers," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2016, pp. 406–407.

[96] T. Takemoto et al., "A 50.6-Gb/s 7.8-mW/Gb/s –7.4-dBm sensitivity optical receiver based on 0.18-μm SiGe BiCMOS technology," in Symp. VLSI Circuits Dig. Tech. Papers, 2016, pp. 1-2.

초 록

최근 데이터 센터나 고성능 컴퓨팅 시스템의 수요가 늘어남에 따라 유선 통신에서 요구되는 대역폭도 기하급수적으로 증가해왔다. 그러나 기존의 구리선을 기반으로 한 방식으로는 이러한 요구를 만족시킬 수 없는데, 그 이유는 구리선은 주파수가 증가함에 따라 손실도 굉장히 증가하기 때문이다. 따라서 장거리 통신에 주로 사용되어 왔던 광 통신을 중거리 내지는 근거리 통신에도 적용해야 할 시점이 되었다. 하지만 기존의 광 통신을 바로 적용하기는 쉽지 않고 여러 문제를 해결해야 하는데, 가장 중요한 점은 기존의 구리 기반 통신 시스템에 비해 높은 에너지 효율과 가격 경쟁력을 갖춰야 한다는 것이다. 이를 해결할 수 있는 기술이 바로 실리콘 포토닉스 (silicon photonics) 이고 이미 예전부터 광범위하게 연구되어 왔다. 실리콘은 광소자와의 호환성은 떨어지지만 효율을 개선할 수 있다면 실리콘 포토닉스는 구리선을 대체할 수 있는 궁극적인 솔루션이 될 것으로 기대가 된다.

본 논문에서는 실리콘 포토닉스의 현 위치에 대해서 요약을 하고 광통신에 대해서 광범위하게 전망을 해본다. 그리고 고속, 저전력 광 수신기 설계를 위한 핵심 회로기술들에 대해서 소개를 한다. 그리고 실제로 구현된 광 수신기에 대해서 기술을 하고, 특히 고성능을 얻기 위해 사용된 새로운 회로기술들에 초점을 맞춘다.

주요어 : 광 통신, 광 수신기, 실리콘 포토닉스, 회로기술

학 번 : 2012-20856